CURRENT MICROBIOLOGY Vol. 44 (2002), pp. 132–135
DOI 10.1007/s00284-001-0063-z

Current Microbiology
An International Journal

Cloning and Molecular Analysis of Poly(3-Hydroxyalkanoate) Biosynthesis Genes in *Pseudomonas aureofaciens*

Tomohiro Nishikawa,[1] Keiko Ogawa,[1] Ryoko Kohda,[1] Wang Zhixiong,[1] Hitoshi Miyasaka,[2] Fusako Umeda,[1] Isamu Maeda,[1] Masaya Kawase,[1] Kiyohito Yagi[1]

[1]Graduate School of Pharmaceutical Sciences, Osaka University, 1-6 Yamada-oka, Suita, Osaka 565-0871, Japan
[2]Kansai Electric Power Company, Inc., Technical Research Center, 3-11-20 Nakoji, Amagasaki, Hyogo 661-0974, Japan

Received: 10 January 2001 / Accepted: 7 June 2001

Abstract. *Pseudomonas aureofaciens* grown on octanoate or gluconate synthesized medium-chain-length polyhydroxyalkanoates (mcl-PHAs). To clone the PHA synthase gene(s) (*phaC*), the genomic library of *P. aureofaciens* was constructed using a cosmid vector. The recombinant cosmids that clone *phaC* were detected by the complementation with a PHA-negative mutant, *P. putida* GPp104. The resulting recombinant cosmid, named pVK6, contained a 13-kbp DNA insert. Genetic analysis of the *pha* locus in pVK6 revealed the presence of six ORFs, genes encoding two PHA synthases, 1 and 2 (*phaC1* and *phaC2*), PHA depolymerase (*phaZ*), two PHA granule-associated proteins (*phaF* and *phaI*), and an unknown protein (*phaD*). The heterologous expression of *pha* genes from *P. aureofaciens* was confirmed. *P. putida* GPp104 regained the ability to accumulate PHA on introduction of pVK6. Wild-type strains *P. oleovorans* and *P. fluorescens*, which were unable to accumulate PHA when grown on gluconate, acquired the ability to accumulate PHA from gluconate when they possessed pVK6.

PHAs are storage polymers accumulated by bacteria during nutrient-limited conditions [2], and are of industrial interest as a material for biodegradable thermoplastic [8]. PHAs are divided into two groups based on their monomer chain length; that is, short-chain-length PHAs (scl-PHAs, C3 to C5) and medium-chain-length PHAs (mcl-PHAs, C6 to C14). scl-PHA production by *Ralstonia eutropha* (formerly *Alcaligenes eutrophus*) has been extensively studied [11]. Polyhydroxybutyrate (PHB), a homopolymer of (R)-3-hydroxybutyrate, is the most investigated scl-PHA. PHB production has already been developed using recombinant *Escherichia coli* carrying *pha* genes from *R. eutropha*, in which high amounts of PHB (more than 90% of the cell dry weight [CDW]) are synthesized and the range of carbon sources used as substrates is being extended [1, 7].

Pseudomonas strains synthesize mcl-PHA [15]. For example, *P. putida* [3, 4] and *P. aeruginosa* [16] can accumulate mcl-PHA not only when grown on alkanes or alkanoic acids but also when grown on unrelated substrates such as gluconate, fructose, and glucose. Comparing PHB with mcl-PHA, mcl-PHA is more useful because PHB is highly crystalline and brittle and the technical applications are limited. Moreover, mcl-PHA contains a variety of monomers and has versatile characteristics. Despite these factors, studies using recombinant *E. coli* as performed for PHB production have not yet been carried out for mcl-PHA production. More research is needed for development of mcl-PHA production. In this study, the results with regards to the composition of PHA, genetic analysis and heterologous expression of the *pha* genes in *P. aureofaciens* are described.

Correspondence to: K. Yagi; *email:* yagi@phs.osaka-u.ac.jp

Materials and Methods

Bacterial strains and cultivation. Bacterial strains used in this study are *P. aureofaciens* IFO3521, *P. putida* GPp104 [5], *P. oleovorans* ATCC29347, and *P. fluorescens* DSM50090. For PHA accumulation, bacterial strains were cultivated in a mineral salt medium containing 0.05% (wt/vol) NH_4Cl as a nitrogen source [10] at 30°C for 72 h. Substrates were added at the final concentrations of 0.5% (wt/vol) for octanoate (0.1% [wt/vol], five times) and 1.5% (wt/vol) for gluconate, glucose, and fructose.

Cloning of PHA synthase gene (*phaC*). The genomic library of *P. aureofaciens* was constructed using the broad-host-range cosmid vec-

Table 1. PHA accumulation in *P. aureofaciens*

Substrate	PHA content (% [wt/wt] of CDW[b])	Composition of PHA (mol%)[a]					
		3HB	3HX	3HO	3HD	3HDD	3HDD:1
Octanoate	45.0	0	12	82	6	0	0
Gluconate	31.7	0	2	12	55	17	14
Glucose	32.2	0	4	15	47	17	17
Fructose	29.1	0	4	18	48	16	13

[a] 3HB, 3-hydroxybutyrate; 3HX, 3-hydroxyhexanoate; 3HO, 3-hydroxyoctanoate; 3HD, 3-hydroxydecanoate; 3HDD, 3-hydroxydodecanoate; 3HDD:1, 3-hydroxydodecenoate.
[b] CDW, cell dry weight.

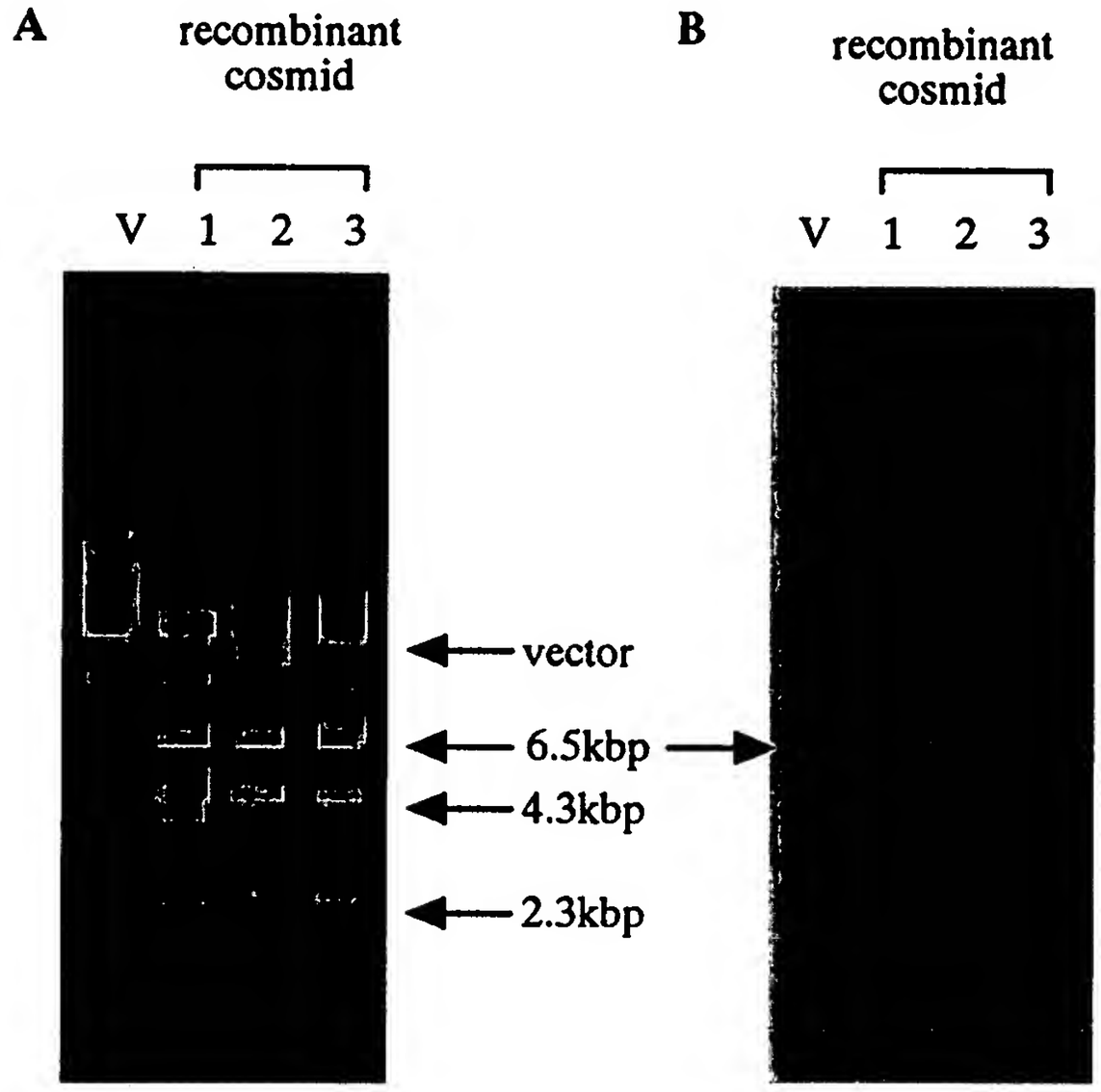

Fig. 1. Detection of *phaC*-encoded fragment. **(A)** agarose gel electrophoresis. Cosmid DNAs digested with *Hin*dIII were charged. **(B)** Southern hybridization. *phaC1* from *P. oleovorans* was used as a probe. Lane V: cosmid vector pVK102; lanes 1, 2, and 3: recombinant cosmids pVK2, pVK4, and pVK6, respectively.

tor, pVK102 [6]. The DNA fragment partially-digested with *Hin*dIII was inserted to the cloning site. The recombinant cosmids that clone *phaC* were detected by the complementation with *P. putida* GPp104, a PHA-negative mutant of *P. putida* KT2442.

DNA sequencing was carried out by the dideoxychain termination method as described by Sanger et al. [13] using the Automatic Sequencer DSQ-2000 L (Shimadzu). The sequencing reaction was performed in accordance with the manual supplied with DNA sequencing kit, the Big Dye Terminator Cycle Sequencing Ready Reaction (Perkin Elmer Biosystems).

Southern hybridization was conducted using *phaC1* from *P. oleovorans* as a probe. The labeled probe was prepared using the ECF Random Prime Labeling Kit (Amersham Pharmacia Biotech) and detection of hybridization signals on membrane was carried out using FluoroImager 595 (Molecular Dynamics).

Analysis of PHA. PHAs were extracted from lyophilized cells with chloroform and then methanolyzed for analysis by gas chromatography [15]. The methyl esters of monomers were assayed using the Gas Chromatograph GC-12 (Shimadzu) and a column packed with Thermon-3000 supported on 60/80 Shincarbon (Shinwakakou) as the stationary phase.

Nucleotide sequence accession number. The nucleotide sequence data are registered in the DDBJ nucleotide sequence database under the accession number AB049413.

Results and Discussion

PHA accumulation in *P. aureofaciens*. *P. aureofaciens* was cultivated on octanoate, gluconate, fructose, or glucose as the sole carbon source. The content and composition of accumulated PHA using these substrates in *P. aureofaciens* are summarized in Table 1. mcl-PHA ac-

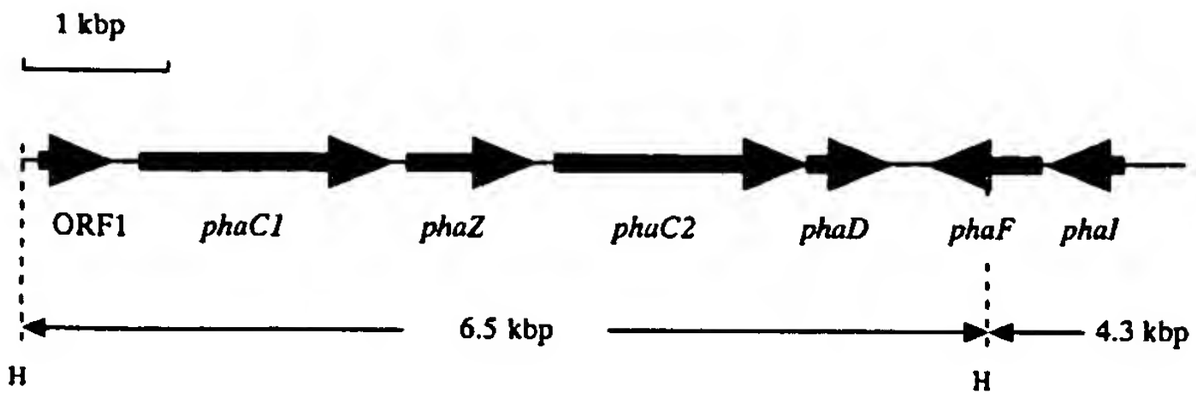

Fig. 2. *pha* locus in *P. aureofaciens*. Proteins (genes): unknown 1 (ORF1), PHA synthase 1 (*phaC1*), PHA depolymerase (*phaZ*), PHA synthase 2 (*phaC2*), unknown 2 (*phaD*), and PHA granule-associated proteins (*phaF* and *phaI*). Arrows show the direction of transcription. H, *Hin*dIII.

Table 2. Heterologous expression of *pha* genes from *P. aureofaciens*

Strain (cosmid carried)	Substrate	PHA content (% [wt/wt] of CDW[b])	Composition of PHA (mol%)[a]					
			3HB	3HX	3HO	3HD	3HDD	3HDD:1
P. putida GPp104 (pVK102)	Gluconate	0	0	0	0	0	0	0
P. putida GPp104 (pVK6)	Gluconate	12.1	0	0	8	54	20	17
P. putida GPp104 (pVK102)	Octanoate	0	0	0	0	0	0	0
P. putida 104 (pVK6)	Octanoate	22.4	0	6	86	5	2	0
P. oleovorans (pVK102)	Gluconate	0	0	0	0	0	0	0
P. oleovorans (pVK6)	Gluconate	7.5	0	0	24	54	15	7
P. oleovorans (pVK102)	Octanoate	40.3	0	8	87	5	0	0
P. oleovorans (pVK6)	Octanoate	48.6	0	7	89	2	0	0

Both strains of *P. putida* GPp104 and *P. oleovorans* each carrying a cosmid were used.
pVK102, broad-host-range cosmid vector; pRK6, pVK102 cloning *pha* genes from *P. aureofaciens*.
[a] 3HB, 3-hydroxybutyrate; 3HX, 3-hydroxyhexanoate; 3HO, 3-hydroxyoctanoate; 3HD, 3-hydroxydecanoate; 3HDD, 3-hydroxydodecanoate; 3HDD:1, 3-hydroxydodecenoate.
[b] CDW, cell dry weight.

cumulated approximately 45% of CDW in the presence of a related carbon source (octanoate), and approximately 30% of the CDW in the presence of unrelated carbon source (gluconate, glucose, or fructose). Accumulated PHA with gluconate, glucose, or fructose as the sole carbon source mainly consisted of 3-hydroxydecanoate (3HD), 3-hydroxydodecanoate (3HDD), and 3-hydroxydodecenoate (3HDD:1). Accumulated PHA in *P. aureofaciens* did not include any 3-hydroxybutyrate (3HB) as a monomer.

Genetic analysis of *pha* genes. Seven recombinant cosmids, which conferred the PHA synthesizing ability to the PHA-negative mutant, *P. putida* GPp104, were isolated. Each recombinant cosmid containing a partially digested *Hin*dIII fragment in the cloning site was completely redigested with *Hin*dIII. The digestion pattern was classified into three types that correspond to recombinant cosmids pVK2, pVK4, and pVK6 (Fig. 1). The recombinant cosmids had 6.5-kbp, 4.3-kbp, and 2.3-kbp *Hin*dIII fragments in common as shown in Fig. 1A. Southern hybridization was carried out using *phaC1* from *P. oleovorans* as a probe. Results showed that *phaC* from *P. aureofaciens* was encoded in the 6.5-kbp DNA fragment (Fig. 1B).

About an 8-kbp area of the 6.5-kbp and flanking 4.3-kbp fragments was sequenced. Genetic analysis revealed the presence of seven ORFs (Fig. 2). The *pha* locus contained genes encoding two PHA synthases, 1 and 2 (*phaC1* and *phaC2*), PHA depolymerase (*phaZ*), an unknown protein (*phaD*), and two PHA granule-associated proteins (*phaF* and *phaI*). A homologous *pha* cluster showing a similar gene organization has been found in mcl-PHA-synthesizing bacteria *P. aeruginosa* [16] and *P. oleovorans* [5, 12].

Heterologous expression of *pha* genes from *P. aureofaciens*. As described in the Materials and Methods section, the genomic library was introduced into *P. putida* GPp104 to detect recombinant cosmids that clone *phaC*. Recombinant, *P. putida* GPp104 cells carrying *phaC*, were selected by their white colony formation due to PHA accumulation on an octanoate plate. Here, PHA accumulation was tested using recombinant *P. putida* GPp104 carrying pVK6, which contained the smallest DNA insert of 13 kbp. The recombinant accumulated PHA when grown on octanoate or gluconate as shown in Table 2. These results show that *phaC1* and/or *phaC2* were/was expressed in *P. putida* GPp104.

P. oleovorans accumulates PHA when grown on

octanoate but not when grown on gluconate [5]. Gluconate-grown cells of *P. oleovorans* carrying pVK6 accumulated PHA and octanoate-grown cells had an increased PHA content as shown in Table 2. These results show that the *pha* genes from *P. aureofaciens* were expressed in *P. oleovorans*. Similar results as obtained in *P. oleovorans* were shown in *P. fluorescens*. *P. fluorescens* do not accumulate any detectable PHA when grown on gluconate [15]. *P. fluorescens* also accumulated PHA at approximately 5% of CDW using gluconate on introduction of pVK6 (data not shown).

Much attention is given to PHA production using renewable sources such as carbon dioxide and industrial and agricultural wastes [1, 7, 9]. Cyanobacteria are promising hosts for PHA production from carbon dioxide because they can grow using carbon dioxide as the sole carbon source. PHA production by cyanobacteria has been studied using both isolates from natural habitat and recombinant cyanobacteria carrying *pha* genes introduced from *R. eutropha* [9, 14]. *E. coli* is another promising host. *E. coli* normally does not accumulate PHA. Furthermore, PHA produced by recombinants with introduced *pha* genes does not decompose since *E. coli* does not carry PHA depolymerase. PHB production has already been successfully carried out using cyanobacteria [9, 14] and *E. coli* [1, 7] as described in the introduction. For practical applications, flexible mcl-PHAs are generally considered to be more useful than brittle PHB. However, studies of mcl-PHA production by cyanobacteria and *E. coli* have hardly progressed. *pha* genes from *P. aureofaciens* will be useful for the mcl-PHA production using renewable sources as substrates.

Literature Cited

1. Ahn WS, Park SJ, Lee SY (2000) Production of poly(3-hydroxybutyrate) by fed-batch culture of recombinant *Escherichia coli* with a highly concentrated whey solution. Appl Environ Microbiol 66:3624–3627
2. Anderson AJ, Dawes EA (1990) Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. Microbiol Rev 54:450–472
3. Huijberts GNM, Eggink G, de Waard P, Huisman GW, Witholt B (1992) *Pseudomonas putida* KT2442 cultivated on glucose accumulates poly(3-hydroxyalkanoates) consisting of saturated and unsaturated monomers. Appl Environ Microbiol 58:536–544
4. Huijberts GNM, de Rijk TC, de Waard P, Eggink G (1994) ^{13}C nuclear magnetic resonance studies of *Pseudomonas putida* fatty acid metabolic route involved in poly(3-hydroxyalkanoate) synthesis. J Bacteriol 176:1661–1666
5. Huisman GW, Wonink E, Mesima R, Kazemier B, Terpstra P, Witholt B (1991) Metabolism of poly(3-hydroxyalkanoates) (PHAs) by *Pseudomonas oleovorans*. J Biol Chem 266:2191–2198
6. Knauf VC, Nester EW (1982) Wide host range cloning vector: a cosmid clone bank of an *Agrobacterium* Ti plasmid. Plasmid 8:45–54
7. Lee SY, Lee KM, Chang HN, Steinbüchel A (1994) Comparison of recombinant *Escherichia coli* strains for synthesis and accumulation of poly-(3-hydroxybutyric acid) and morphological changes. Biotechnol Bioeng 55:1337–1347
8. Madison LL, Huisman GW (1999) Metabolic engineering of poly(3-hydroxyalkanoate): from DNA to plastic. Microbiol Mol Biol Rev 63:21–53
9. Miyake M, Takase K, Narato M, Khatipov E, Schnackenberg J, Shirai M, Kurane R, Asada Y (2000) Polyhydroxybutyrate production from carbon dioxide by cyanobacteria. Appl Biochem Biotechnol 84–86:991–1002
10. Ohi K, Takada N, Komemushi S, Okazaki M, Miura Y (1979) A new species of hydrogen-utilizing bacterium. J Gen Appl Microbiol 25:53–58
11. Peoples OP, Sinskey AJ (1989) Poly-β-hydroxybutyrate (PHB) biosynthesis in *Alcaligenes eutrophus* H16. Identification and characterization of the PHB polymerase gene (*phbC*). J Biol Chem 264:15298–15303
12. Prieto MA, Buhler B, Jung K, Witholt B, Kessler B (1999) PhaF, a polyhydroxyalkanoate-granule-associated protein of *Pseudomonas oleovorans* GPo1 involved in the regulatory expression system for pha genes. J Bacteriol 181:858–868
13. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors Proc Natl Acad Sci USA 74:5463–5467
14. Taroncher-Oldenburg G, Nishina K, Stephanopoulos G (2000) Identification and analysis of the polyhydroxyalkanoate-specific β-ketothiolase and acetoacetyl coenzyme A reductase genes in the cyanobacterium *Synechocystis* sp. strain PCC6803. Appl Environ Microbiol 66:4440–4448
15. Timm A, Steinbüchel A (1990) Formation of polyesters consisting of medium-chain-length 3-hydroxyalkanoic acids from gluconate by *Pseudomonas aeruginosa* and other fluorescent pseudomonads. Appl Environ Microbiol 56:3360–3367
16. Timm A, Steinbüchel A (1992) Cloning of molecular analysis of the poly(3-hydroxyalkanoic acid) gene locus of *Pseudomonas aeruginosa* PAO1. Eur J Biochem 209:15–30

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 267, No. 35, Issue of December 15, pp. 25513-25520, 1992
Printed in U.S.A.

Cloning, Sequencing, and Expression of the *fadD* Gene of *Escherichia coli* Encoding Acyl Coenzyme A Synthetase*

(Received for publication, May 1, 1992)

Paul N. Black‡, Concetta C. DiRusso, Amy K. Metzger, and Tamra L. Heimert

From the Department of Biochemistry, College of Medicine, University of Tennessee, Memphis, Tennessee 38163

In the enteric bacterium, *Escherichia coli*, acyl coenzyme A synthetase (fatty acid:CoA ligase (AMP-forming) EC 6.2.1.3) activates exogenous long-chain fatty acids concomitant with their transport across the inner membrane into metabolically active CoA thioesters. These compounds serve as substrates for acyl-CoA dehydrogenase in the first step in the process of β-oxidation. The acyl-CoA synthetase structural gene, *fadD*, has been identified on clone 6D1 of the Kohara *E. coli* gene library and by a process of subcloning and complementation analyses shown to be contained on a 2.2-kilobase *NcoI-ClaI* fragment of genomic DNA. The polypeptide encoded within this DNA fragment was identified following T7 RNA polymerase-dependent induction and estimated to be M_r = 62,000 using SDS-polyacrylamide gel electrophoresis. The N-terminal amino acid sequence of acyl-CoA synthetase was determined by automated sequencing to be Met-Lys-Lys-Val-Trp-Leu-Asn-Arg-Tyr-Pro. Sequence analysis of the 2.2-kilobase *NcoI-ClaI* fragment revealed a single open reading frame encoding these amino acids as the first 10 residues of a protein with a molecular weight of 62,028. The initiation codon for methionine was TTG. Primer extension of total *in vivo* mRNA from two *fadD*-specific oligonucleotides defined the transcriptional start at an adenine residue 60 base pairs upstream from the predicted translational start site. Two FadR operator sites of the *fadD* gene were identified at positions −13 to −29 (O_{D1}) and positions −99 to −115 (O_{D2}) by DNase I footprinting. Comparisons of the predicted amino acid sequence of the *E. coli* acyl-CoA synthetase to the deduced amino acid sequences of the rat and yeast acyl-CoA synthetases and the firefly luciferase demonstrated that these enzymes shared a significant degree of similarity. Based on the similar reaction mechanisms of these four enzymes, this similarity may define a region required for the same function.

Exogenous long-chain fatty acids (C_{12}-C_{18}) represent an important class of hydrophobic compounds that can serve as a sole carbon and energy source to support the growth of the enteric bacterium *Escherichia coli*. The acquisition of these nutrients from the environment prior to metabolic utilization by cyclic β-oxidation in *E. coli* occurs by an energy-dependent, protein-mediated process. For long-chain fatty acids destined for β-oxidation, this process minimally requires the products of the *fadL* and *fadD* genes. The *fadL* gene encodes an outer membrane-bound protein (FadL) that binds exogenous long-chain fatty acids with a relatively high affinity and by some unknown mechanism transfers these compounds across the outer membrane (1–4). The *fadD* gene encodes the inner membrane-associated acyl coenzyme A synthetase (acyl-CoA synthetase (fatty acid:coenzyme A ligase (AMP-forming), EC 6.2.1.3)) (5). This enzyme catalyzes the esterification of fatty acids into metabolically active CoA thioesters concomitant with transport. The mechanisms that govern the transfer of long-chain fatty acids from FadL across the periplasmic space and the inner membrane to the acyl-CoA synthetase remain largely undefined. There is some evidence for an oleic acid binding protein in the inner membrane that has been postulated to be a H^+/long-chain fatty acid co-transporter (6, 7).

Acyl-CoA synthetases catalyze the formation of fatty acyl-CoA by a two-step mechanism that proceeds through the pyrophosphorolysis of ATP (8).

$$\text{fatty acid} + \text{ATP} \xrightleftharpoons{Mg^{2+}} [\text{fatty acid} - \text{AMP}] + PP_i$$

STRUCTURE I

$$[\text{fatty acid} - \text{AMP}] + \text{CoASH} \rightarrow \text{fatty acid} - \text{SCoA} + \text{AMP}$$

STRUCTURE II

E. coli contains a single acyl-CoA synthetase which has been purified to homogeneity (9, 10). This enzyme has broad chain-length specificity giving V_{max} values ranging from 2632 nmol/min/mg of protein for lauric acid (C_{12}) to 135 nmol/min/mg of protein for hexanoate (C_6) (9). Maximal activities associated with this enzyme are found with fatty acids ranging in length between C_{12} and $C_{18:1}$ (9). Overath and colleagues (11) proposed that acyl-CoA synthetase of *E. coli* was required for long-chain fatty acid transport and coined the term vectorial acylation to describe the role of this enzyme. Although the precise role of this enzyme in fatty acid transport is not well defined, it is clear that it plays a pivotal role in this process by catalyzing the formation of metabolically active CoA thioesters for subsequent degradation or incorporation into phospholipids (5, 9, 11).

The structural gene for acyl-CoA synthetase (*fadD*) was identified by Overath *et al.* (5) who mapped this locus to the 40-min region of the *E. coli* chromosome. In this pioneering work, this enzyme was proposed to be partially membrane-associated and was shown to activate both mono- and poly-

* This work was supported by National Science Foundation Grant 9104646 and by American Heart Association Grant 901063 (to P. N. B.) and by National Institutes of Health Grant GM38104 (to C. C. D.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) L02649.

‡ To whom correspondence and reprint requests should be addressed: Dept. of Biochemistry, Suite G01, Molecular Sciences Bldg., 858 Madison Ave., University of Tennessee College of Medicine, Memphis, TN 38163. Tel.: 901-528-6576; Fax: 901-528-7360.

unsaturated fatty acids. Acyl-CoA synthetase from *E. coli* has been estimated to have a native molecular weight of 120,000 based on elution profiles of the purified enzyme on a G-200 column (5). Kameda and Nunn (9) estimated a monomeric molecular weight of 47,000 and proposed that the enzyme is a dimer. Acyl-CoA synthetase activity is induced by oleate, but to a lower relative level, when compared to the levels of induction documented for three other enzymes required for long-chain fatty acid degradation (β-hydroxyacyl-CoA dehydrogenase, enoyl-CoA hydratase, and β-ketothiolase) (5). The *fadD* gene, like the *fadBA*, *fadE*, and *fadL* genes, is part of the fatty acid degradative regulon under the control of the transcriptional regulator FadR (11–13).

The present work describes the cloning, sequencing, and expression of the acyl-CoA synthetase structural gene (*fadD*) of *E. coli*. This work stems from our goal to define the underlying biochemical mechanisms that govern long-chain fatty acid transport in enteric bacteria prior to metabolic utilization.

EXPERIMENTAL PROCEDURES

Bacterial Strains—The *E. coli* strain JM103 [Δ(*lac pro*) *thi strA endA sbcB hsdR*(F′ *traD36 proAB lacI*^{q} *ZΔM15*)] was used for the propagation of M13 derivatives. For routine plasmid propagation and the generation of the *fadD88* strain PN235, C600 (F-*thi-1 leuB6 lacY1 tonA21 supE44*) was used. Strains RS3010 (*fadR*) and LS6928 (*fadD88 zea*::Tn*10 fadR*) have been described elsewhere (14, 15). The *fadD88* strain PN235 was generated by P1 transduction of strain C600 using a phage stock grown on strain LS6928. Bacterial cultures were grown at 37 °C in a Lab Line gyratory shaker in 2YT (16), Luria broth (LB; 16), or Tryptone broth (TB; 17). When minimal medium was required, medium E supplemented with vitamin B_1 (17) was used. Carbon sources, sterilized separately, were added to final concentrations of 25 mM glucose, 25 mM potassium acetate, 5 mM decanoate, or 5 mM oleate. As required, amino acids were added to a final concentration of 0.01%. When required to maintain plasmids, antibiotics were added to 100 μg/ml ampicillin, 40 μg/ml kanamycin, 10 μg/ml tetracycline, and 40 μg/ml chloramphenicol. Growth of bacterial cultures was routinely monitored using a Klett-Summerson colorimeter equipped with a blue filter.

Identification of the fadD Gene in the Kohara Gene Library—The complete miniset of λ clones of the Kohara library (18) were graciously provided by Dr. Y. Kohara (DNA Research Center, National Institute of Genetics, Mishima 411, Japan). Eleven clones representing the 40-min region of the *E. coli* chromosome (5E12, 4B8, 12H7, 3E12, 9F2, 7F2, 6D1, 12B3, 15D5, 19H3, and 12C7) were propagated on the bacterial strain NM621 as previously described (19). Lysates were used to infect the λCI857 lysogen derived from strain PN235 at a multiplicity of infection of 1 as described by Miller (17). Following absorption, 1 ml of LB was added and the cells were allowed to recover for 30 min at 30 °C. Following recovery, the cells were pelleted by centrifugation, resuspended in the original volume of Medium E, plated on oleate minimal agar plates, and incubated at 30 °C for 72 h. At 72 h, colonies that were able to grow on oleate as a sole carbon and energy source were identified in cells infected with phage DNA from clones 7F2 and 6D1. These two clones were confirmed to restore the ability of the *fadD* strain PN235 to grow on oleate (Ole$^+$) following a second round of lysogenic complementation. Clones 7F2 and 6D1 from the Kohara library were propagated in strain NM621 for DNA isolation on TB agarose plates (24). Plaques giving nearly confluent lysis were visible 12–14 h later at which time the plates were flooded with 3 ml of 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM $MgSO_4$, 0.01% gelatin (SM). λ-DNA was isolated using the hexadecyltrimethylammonium bromide (CTAB) method (20). Briefly, 20 ml of each phage stock (4×10^{10} plaque-forming units/ml) was incubated with DNase I at 20 μg/ml for 5 min and then clarified by centrifugation (10,000 × *g* for 20 min). The supernatants were transferred to a new tube, and 10 ml of DEAE-cellulose slurry (80% in SM) was added and incubated at room temperature for 30 min in an angled rotator. The DEAE-cellulose was pelleted by centrifugation, the supernatants were transferred to a new tube, and EDTA was added to 20 mM and Tris-HCl, pH 8.0, was added to 100 mM. Proteinase K was added to the mixture to a final concentration of 50 μg/ml which was then heated at 45 °C for 15 min. Hexadecyltrimethylammonium bromide was added to a final concentration of 0.5% and incubated at 68 °C for 4 min. The samples were cooled on ice and centrifuged 10,000 × *g* for 30 min. The DNA pellets were dissolved in 6 ml of 1.2 M NaCl and DNA precipitated by the addition of 2 volumes of absolute ethanol. The final DNA samples were resuspended in 200 μl of 10 mM Tris-HCl, pH 7.5, 5 mM EDTA (TE), analyzed by agarose gel electrophoresis, and used in subcloning experiments as described below.

Cloning and Sequencing—DNA from clone 6D1 was restricted with *Cla*I or *Hin*dIII, ligated into the plasmid vector pACYC177 (21), and transformed into the *fadD88* strain PN235. Restriction, ligation, plasmid isolation, and transformation procedures have been described previously (22, 23). Ole$^+$ transformants were identified in both sets of ligation mixtures. Analysis of restriction patterns generated revealed that restriction fragments from the Ole$^+$ transformants from the *Cla*I ligation mixture and the *Hin*dIII ligation mixture overlapped. As the insert from the *Cla*I digest was shown to be smaller (3.4 kb),[1] this plasmid, designated pN300, was used for all further study. Restriction fragments from pN300 were subcloned into either pACYC177 or pACYC184 (21) as described under "Results," yielding the plasmids and complementation patterns illustrated in Fig. 1*B*. The sequencing strategy of the *Cla*I insert from pN300 is illustrated in Fig. 1*C*. The series of M13 clones were sequenced using either the *lacZ*-specific upstream primer (5′-GTTTTCCCAGTCACGAC-3′), the M13 universal primer (5′-GTAAAACGACGGCCAGT-3′), or with *fadD*-specific oligonucleotides by the dideoxy chain-terminating method of Sanger *et al.* (24) using Sequenase (v 2.0; U. S. Biochemicals). As shown in Fig. 1*C*, pN300 was sequenced across the *Sal*I and *Hin*dIII restriction sites using *fadD*-specific oligonucleotides to ensure proper alignment between these three fragments of DNA. Sequencing reactions were resolved on a standard 8% polyacrylamide gel (3). All oligonucleotides used in this study were synthesized on a Pharmacia LKB Biotechnology Inc. Gene Assembler Plus.

Analysis of Acyl-CoA Synthetase Activity—Bacteria (wild-type and *fadD* strains containing the collection of *fadD*$^+$ and *fadD* clones) were grown to midlog phase (6×10^8 cells/ml) in TB or TB supplemented with 5 mM oleate and 0.5% Brij 58 (TBO) and with antibiotics as required. Cells were harvested by centrifugation, washed twice with Medium E, resuspended to a density of 1.2×10^9 cells/ml in 10 mM Tris-HCl, pH 7.5, and lysed by three cycles of sonication at 0 °C. Acyl-CoA synthetase activities were determined in sonicated cell extracts as described by Kameda and Nunn (9). The reaction mixtures contained 200 mM Tris-HCl, pH 7.5, 2.5 mM ATP, 8 mM $MgCl_2$, 2 mM EDTA, 20 mM NaF, 0.1% Triton X-100, 10 μM [^{3}H]oleate, 0.5 mM coenzyme A, and cell extract in a total volume of 0.5 ml. The reactions were initiated with the addition of coenzyme A, incubated at 35 °C for 10 min, and terminated by the addition of 2.5 ml of isopropyl alcohol:*n*-heptane:1 M H_2SO_4 (40:10:1). The radioactive oleic acid was removed by organic extraction using *n*-heptane (9). Oleoyl-CoA formed during the reaction remained in the aqueous fraction and was quantified by scintillation counting. Protein concentrations in the enzyme extracts were determined using the Bradford assay and bovine serum albumin as a standard (25). The values presented represent the average from at least three independent experiments.

Overexpression of Acyl-CoA Synthetase—The 3.4-kb *Cla*I fragment (*fadD*$^+$) from pN300 was gel-purified and ligated with a *Cla*I to *Bam*HI linker. Linkers were purified and phosphorylated using bacterial alkaline phosphatase prior to restriction (22). Following ligation of the linkers, the fragment was restricted with *Bam*HI, repurified, and ligated into the *Bam*HI site of the expression plasmid pCD130. pCD130 is derived from pT7-5 and contains the *fadR* gene in the orientation opposite to the T7 promoter to maintain stability of *fadD*$^+$ (26). Both orientations of the *fadD* gene were obtained yielding plasmid pN321 and pN324. Plasmid pN324 contained *fadD*$^+$ under the T7 promoter while pN321 had *fadD*$^+$ in the orientation opposite to the T7 promoter. The plasmids pN321 and pN324 were transformed into strain BL21 (DE3)(plysS) and expressed following induction with isopropyl-1-thio-β-D-galactopyranoside (27). Following induction and labeling with [^{35}S]methionine, cells were harvested, resuspended in SDS sample buffer, boiled, and resolved on a 12% SDS-polyacrylamide gel using the Laemmli buffer system (28). Following electrophoresis, the gels were dried and subjected to autoradiography for 2–24 h.

Partial Purification of Acyl-CoA Synthetase and N-terminal Amino Acid Sequencing—Acyl-CoA synthetase was partially purified from a

[1] The abbreviation used is: kb, kilobase pair(s).

500-ml culture of strain BL21(pLysS) harboring the *fadD*$^+$ expression plasmid pN324 after induction. Following induction with isopropyl β-D-thiogalactopyranoside, cultures were grown for an additional 2 h and cells were harvested by centrifugation. The cell pellets were washed twice in minimal medium E, resuspended in 50 mM potassium phosphate, pH 8.0, and disrupted by three cycles of sonication at 0 °C. The sonicated extract was clarified by centrifugation (12,000 × *g* for 15 min). The supernatant was centrifuged at 60,000 × *g* for 2 h, and the membrane pellet was discarded. Acyl-CoA synthetase was partially purified from the supernatant by batch DEAE-cellulose chromatography and by ammonium sulfate fractionation using the conditions described by Overath *et al.* (5) and Kameda *et al.* (10). The final protein sample was subjected to preparative electrophoresis on a 12% SDS-polyacrylamide gel and electrophoretically transferred to a prewetted polyvinylidene difluoride (Immobilon Transfer, Whatman) in 10 mM 3-(cyclohexylamino)-1-propanesulfonic acid, pH 11.0, 10% methanol (29). Following electrophoretic transfer, the position of the acyl-CoA synthetase was identified by staining the polyvinylidene difluoride membrane briefly with Ponceau red. The strip containing acyl-CoA synthetase was excised, extensively washed with high performance liquid chromatography grade water, dried at room temperature, and stored at −70 °C. The N-terminal amino acid sequence from this sample was determined using an Applied BioSystems 470A gas phase protein sequenator equipped with an Applied BioSystems Model 120A in-line detector for phenylthiohydantoin-derived amino acids from each cycle of Edman degradation at the Harvard University MicroChemistry Facility.

Mapping the Transcription Initiation Site of the fadD Gene—The transcriptional start of the *fadD* gene was identified by primer extension of two *fadD*-specific oligonucleotides, 5′-GATAACGGTTAAGCCAAACC-3′ and 5′-CTACCAGAGATTGATAACGG-3′ (corresponding to nucleotides 315–334 and 366–385, respectively; see Fig. 4), hybridized to *in vivo*-synthesized mRNA by reverse transcriptase (3). Total *in vivo*-synthesized RNA was isolated from a 100-ml midlog culture of strain RS3010 (*fadR*) or K12 grown in TB (3). The *fadD*-specific oligonucleotides for primer extension were 5′ end-labeled using [γ-^{32}P]ATP and polynucleotide kinase as described by Maniatis *et al.* (22).

Identification of the FadR Binding Site by DNase I Footprinting—The 353-base pair *Sau*3A fragment containing the *fadD* promoter was gel-purified and ligated into the *Bam*HI site of M13mp19, M13mp18, and pUC18 thereby generating clones MD21 (to sequence the top strand of the *fadD* promoter-containing fragment), MD20 (to sequence the bottom strand of the *fadD* promoter-containing fragment), and pN330, respectively. Plasmid pN330 was purified and used as a source of DNA for gel shifts and DNase I footprinting assays. Protein-DNA gel retention assays (gel shifts) and DNase I footprinting were carried out essentially as described by DiRusso *et al.* (13) using purified FadR. The concentrations of FadR used in these experiments are given in the appropriate figure legends. For gel shifts, the 414-base pair *Eco*RI-*Hin*dIII fragment from pN330 was gel-purified and labeled with [α-^{32}P]dATP using the Klenow fragment of DNA polymerase (22). The binding of FadR to an [α-^{32}P]dATP-labeled fragment containing the *fadB* operator was used as an internal control as these parameters were previously well defined (13). FadR binding was estimated as the conversion of the fast mobility complex (unbound DNA) to the slow mobility complex (FadR-bound). Quantitation was carried out with a BioImage computer-assisted analysis system (MilliGen/Biosearch). For DNase I footprinting, pN330 was restricted with *Eco*RI (top strand) or *Hin*dIII (bottom strand) and *Pvu*II (cleaves only within the vector), and the promoter-containing fragments were gel-purified. Following purification, the *Hin*dIII-*Pvu*II and *Eco*RI-*Pvu*II fragments were 5′ end-labeled with [γ-^{32}P]ATP using polynucleotide kinase and restricted with *Eco*RI and *Hin*dIII, respectively. The appropriate ^{32}P-labeled *Eco*RI-*Hin*dIII (top strand of the *fadD* promoter labeled) or *Hin*dIII-*Eco*RI (bottom strand of the *fadD* promoter labeled) fragments were gel-purified and used for DNase I footprinting using the conditions described by DiRusso *et al.* (13). The concentrations of FadR used in these experiments are given in the appropriate figure legends. Clones MD21 and MD20 were sequenced using the *Eco*RI-specific primer and the *Hin*dIII-specific primer, respectively, and sequencing reactions were run adjacent to the DNase I footprint reactions on a 6% standard sequencing gel to accurately position the FadR binding site(s) within the *fadD* promoter (17).

The analyses of the DNA sequence of the *fadD* gene and the amino acid sequences of acyl-CoA synthetase, yeast acyl-CoA synthetase, rat acyl-CoA synthetase, and firefly luciferase were done using the Wisconsin Genetics Computer Group programs (35) and DNA Inspector II (TextCo Inc., West Lebanon, NH).

Materials—Reagents and enzymes used for sequencing, transcription mapping, and restriction were purchased from U. S. Biochemicals, Bethesda Research Laboratories, and New England Biolabs. Reagents used for oligonucleotide synthesis were purchased from ABN/Biogenex and Pharmacia. [α-^{35}S]dATP, [α-^{32}P]dATP, [γ-^{32}P]ATP, [^{35}S]methionine, and [^{3}H]oleate were obtained from Du Pont-New England Nuclear. Antibiotics and other supplements for bacterial growth were purchased from Difco and Sigma. All other chemicals were obtained from standard suppliers and were of reagent grade.

RESULTS

Cloning the fadD Gene—The *fadD* gene was mapped by Overath *et al.* (5) to the 40-min region of the *E. coli* chromosome (5). Phage from 11 clones (5E12, 4B8, 12H7, 3E12, 9F2, 7F2, 6D1, 12B3, 15D5, 19H3, and 20H4) of the λ gene library generated by Kohara *et al.* (18) were transduced into the λCI857 lysogen of the *fadD* strain PN235 (Fig. 1*A*) for complementation analysis. DNA from clones 7F2 and 6D1 was able to complement the *fadD88* defect. λ-DNA was purified from clone 6D1 and restricted with *Cla*I, and fragments were ligated into *Cla*I-restricted pACYC177. Plasmid DNA from one of the transformants that complemented the *fadD88* mutation (acquired the ability to grow on the long-chain fatty acid oleate (Ole$^+$)) was purified and designated pN300 (Fig. 1*B*). This plasmid contained a 3.4-kb *Cla*I fragment of ge-

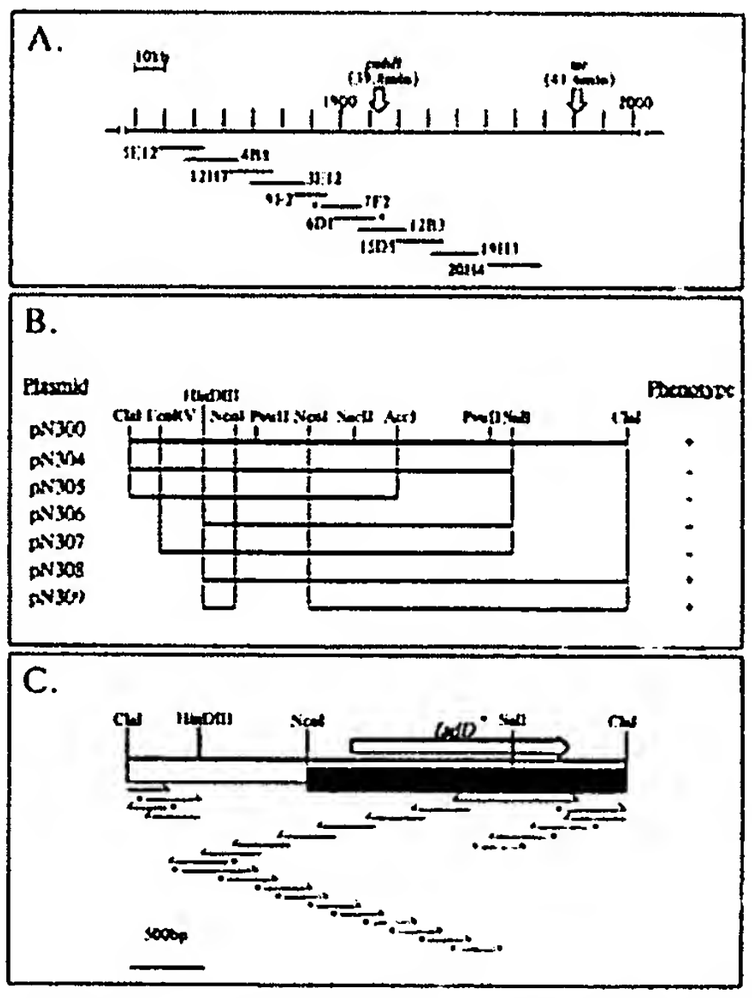

FIG. 1. **Cloning and sequencing the *fadD*$^+$ gene of *E. coli*.** *A*, the *top line* represents the *E. coli* chromosome with the *numbers* representing the position in kilobases. The genetic positions of *pabB* and *tar* are given as reference points. The *small lines* represent the chromosomal DNA in the different isolates (numbered according to serial number) of the Kohara gene library (23). Clones 7F2 and 6D1 are *starred* to indicate that they were able to complement the *fadD88* defect in strain PN235. *B*, restriction map of pN300 and *fadD* and *fadD*$^+$ subclones. The complementation patterns are shown at the *right*, + refers to growth on 5 mM oleate, and − refers to no growth on 5 mM oleate. *C*, sequencing strategy of the *fadD*$^+$ gene. *Arrows* represent M13 subclones of the *fadD* gene including those isolates generated using *Exo*III. These clones were sequenced using either the universal primer or the *lacZ*-specific primer. *Arrows* preceded by an *asterisk* (*) indicate the direction and extent of sequencing using either M13-derived clones or pN130 sequenced using *fadD*-specific oligonucleotides. The *shaded region* (*Nco*I-*Cla*I) represents the sequence presented in Fig. 4.

nomic DNA. A series of subclones of pN300 were constructed to delineate the end points of the *fadD* gene for further studies using DNA sequencing (Fig. 1*B*). One of these subclones, pN308, contained a 2.7-kb *Hin*dIII-*Cla*I fragment which complemented *fadD88*. When a small (500-base pair) *Nco*I fragment was removed from pN308 to generate pN309, the smallest subclone complementing the *fadD88* defect was generated.

Acyl-CoA Synthetase Expression—Acyl-CoA synthetase activities were monitored in the wild-type strain K12, the *fadR* strain RS3010, and the *fadR fadD* strain LS6928 harboring the *fadD*$^+$ and *fadD* plasmids shown in Fig. 1*B*. As shown in Fig. 2*A*, acyl-CoA synthetase activity was inducible 2-fold in the presence of the long-chain fatty acid oleate in the prototrophic strain K12. Acyl-CoA synthetase activities in the *fadR* strain RS3010 grown under both conditions were comparable to the levels found in *E. coli* K-12 grown in the presence of oleate. As expected, no acyl-CoA synthetase activity was observed in the *fadD fadR* strain LS6928. The *fadR fadD* strain LS6928 harboring the *fadD* and *fadD*$^+$ plasmids pN300, pN104, pN305, pN306, pN307, pN308, or pN309 had acyl-CoA synthetase activities that reflected their complementation patterns (Fig. 2*B*).

Identification of the fadD Gene Product and N-terminal Amino Acid Sequence Analysis of Acyl-CoA Synthetase—The 3.4-kb *Cla*I fragment from pN300, containing the entire *fadD* gene, was isolated, ligated with a *Cla*I to *Bam*HI linker, and cloned into the T7 expression plasmid pCD130 (13) to yield the plasmids pN321 and pN324 (Fig. 3*A*). A protein with an M_r of 62,000 was identified by SDS-polyacrylamide gel electrophoresis in extracts of cells harboring pN324 following induction which was presumed to be acyl-CoA synthetase (Fig. 3*B*). A second protein that was poorly produced relative to acyl-CoA synthetase with an M_r of 22,000 was also identified in these extracts which, based on the sequence data described below, was presumed to be distinct from the *fadD* gene. Neither protein was produced in cells harboring pN321 (*fadD*$^+$ in the reverse orientation to the T7 promoter) or pCD130 (plasmid vector)(data not shown). A polypeptide with an M_r of 15,000 was produced from both constructs as well as

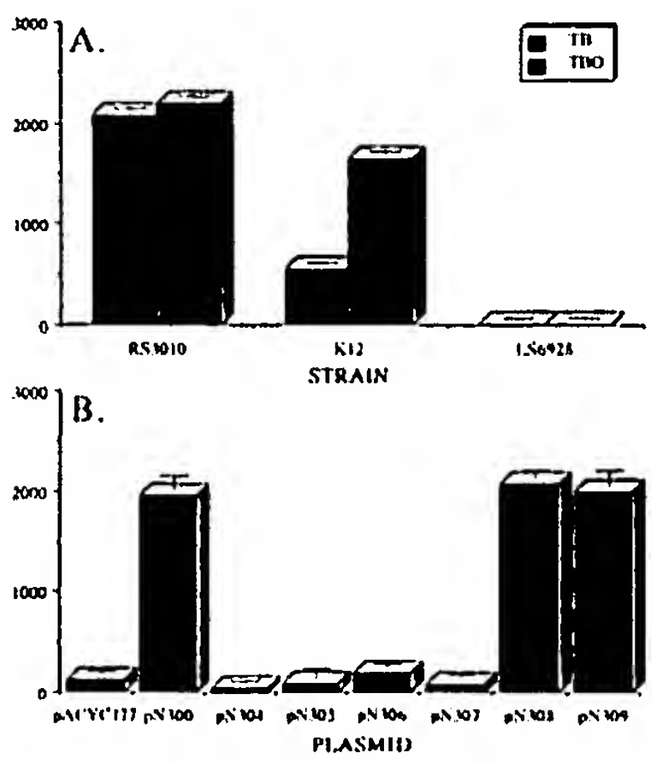

FIG. 2. **Acyl-CoA synthetase activities.** *A*, induction patterns of acyl-CoA synthetase activities in the prototrophic strain K-12, the *fadR* strain RS3010, and the *fadR fadD* strain LS6928 following growth in TB or TB containing 5 mM oleate. *B*, acyl-CoA synthetase activities in the *fadD fadR* strain LS6928 harboring the collection of *fadD*$^+$ and *fadD* clones illustrated in Fig. 1*B* following growth in TB containing 5 mM oleate. The *scale* on the *left* indicates acyl-CoA synthetase activity expressed in picomoles of oleoyl-CoA formed/min/mg of protein. The *error bars* indicate the standard error of the mean of three independent experiments.

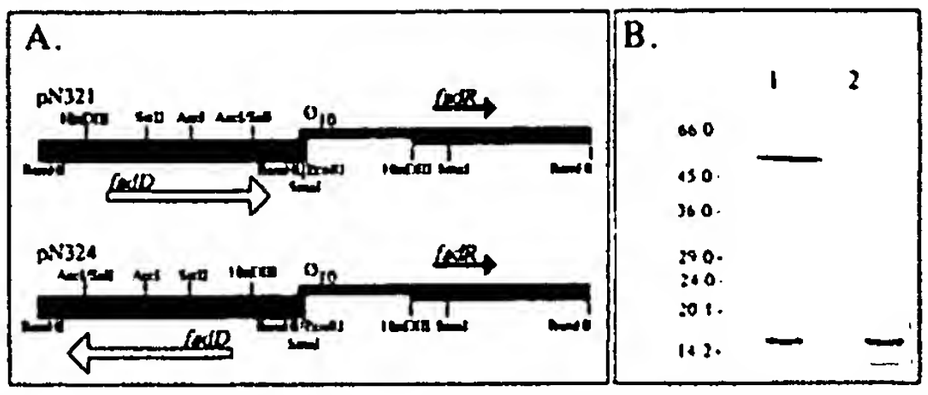

FIG. 3. **Overexpression of acyl-CoA synthetase.** *A*, plasmid constructions harboring the *fadD*$^+$ gene in the (+) orientation (pN324) and the (−) orientation (pN321) relative to the T7 promoter (ϕ_{10}) in plasmid pCD130. *B*, autoradiograph of acyl-CoA synthetase expression following induction with isopropyl-1-thio-β-D-galactopyranoside, labeling with [^{35}S]methionine, and SDS-polyacrylamide gel electrophoresis; *lane 1*, pN324; *lane 2*, pN321; molecular weights ($\times 10^{-3}$) are indicated to the *left*.

the plasmid vector and thus presumed to be vector-specific. Use of this expression system defined the direction of transcription of the *fadD* gene from the *Nco*I site toward the *Cla*I site as shown in Fig. 1*C*. This directional alignment of transcription was confirmed by the DNA sequence of the *fadD* gene.

The M_r = 62,000 polypeptide assumed to be acyl-CoA synthetase was partially purified from BL21(plysS)(pN324) following induction as described under "Experimental Procedures" and subjected to automated N-terminal amino acid sequence analysis as described under "Experimental Procedures." The N-terminal amino acid sequence was shown to be Met-Lys-Lys-Val-Trp-Leu-Asn-Arg-Tyr-Pro.

Sequence of the fadD Gene—The entire 3.4-kb *Cla*I fragment from pN300 was sequenced as shown in Fig. 1*C*. The *fadD* gene was shown by complementation to be localized on a 2.2-kb *Nco*I-*Cla*I fragment. Sequence analysis of this fragment of DNA as shown in Fig. 4 (2230 base pairs) revealed a single open reading frame beginning with ATG at nucleotide 241 encoding a polypeptide consisting of 580 amino acid residues with a molecular weight of 64,406. This open reading frame did not encode the N-terminal amino acid sequence defined from the purified acyl-CoA synthetase (see above). Furthermore, this ATG was five nucleotides upstream from the adenine residue defined as the transcriptional initiation site by primer extension (see below). Careful analysis of the reading frame revealed that the amino acid sequence Leu-Lys-Lys-Val-Trp-Leu-Asn-Arg-Tyr-Pro, starting at nucleotide 307, had 9 of the 10 amino acid residues defined for the purified protein. The notable difference was in the N-terminal amino acid. The DNA sequence predicted a leucine (TTG) while the protein sequence defined a methionine (ATG). We propose that the TTG beginning at nucleotide 307 represented the initiation codon and encodes a methionine residue (thus giving the predicted N-terminal amino acid sequence Met-Lys-Lys-Val-Trp-Leu-Asn-Arg-Tyr-Pro). This proposal was based on three lines of evidence. First, the data obtained from the N-terminal amino acid sequence of the acyl-CoA synthetase clearly indicated the presence of a methionine at this position. Second, the protein sequence beginning at nucleotide 240 did not result in an amino acid sequence that would indicate that this protein was post-translationally modified (*i.e.* cleavage of a signal peptide and/or modification of the N-terminal amino acid). Third, UUG can act as an alternative initiation codon (30). Assuming that the TTG at nucleotide 307 encodes the initiation methionine, the predicted size of the acyl-CoA synthetase from our expression of pN324 (62,000) was in close agreement with that predicted from the DNA sequence (62,028). The coding sequence for the M_r =

```
1    GATCCATGGTTTTATGGCGGTCGTGGCTGGCCCTATGGCTACGGCCGATGGGGCTGGTAT   60
61   AATCCCGGCCCCCGCGAGAGTACAAACAGTTGTAACTGAATAATTGCTTGTTTTTAAAGAA  120
                    OD2
121  AAAGAAACAGCGGCTGGTGCGCTGTTTCTGCATTCTTACGGTAAAGATAAAAATAAATAG  180
                        -35          OD1    -10
181  TGACGCGCTTCGCAACCTTTTCGTTGGGTAATTATCAAGCTGGTATGATGAGTTAATATT  240
          +1
241  ATGTTAACGGCATGTATATCATTTGGGGTTGCGATGACGACGAACACGCATTTTAGAGGT  300
301  GAAGAATTGAAGAAGGTTTGGCTTAACCGTTATCCCCCGGACGTTCCGACGGAGATCAAC  360
       MetLysLysValTrpLeuAsnArgTyrProAlaAspValProThrGluIleAsn
361  CCTGACCGTTATCAATCTCTGGTAGATATGTTTGAGCAGTCGGTCGGCGCGCTACGCCGA  420
       ProAspArgTyrGlnSerLeuValAspMetPheGluGlnSerValGlyAlaLeuArgArg
421  TCAACCTGCGTTTGTGAATATGGGCGAGGTAATGACCTTCGCAAGCTGGAAGAACGCAGT  480
       SerThrCysValCysGluTyrGlyGlyGlyAsnAspLeuArgLysLeuGluGluArgSer
481  CGCGCGTTTGCCGCTTATTTGCAACAAGGGTTGGGGCTGAAGAAAGGCGATCCGCGTTGCG  540
       ArgAlaPheAlaAlaTyrLeuGlnGlnGlyLeuGlyLeuLysLysGlyAspArgValAla
541  TTGATCATGCCTAATTTATTGCAATATCCGGTGGCGCTGTTTGGCATTTTGCGTGCCGGG  600
       LeuMetMetProAsnLeuLeuGlnTyrProValAlaLeuPheGlyIleLeuArgAlaGly
601  ATGATCGTCGTAAACCGTTAACCCGTTGTATACCCCGCGTGAGCTTGAGCATCAGCTTAAC  660
       MetIleValValAsnValAsnProLeuTyrThrProArgGluLeuGluHisGlnLeuAsn
661  GATAGCGGCGCATCGGCGATTGTTATCGTGTCTAACTTTGCTCACACACTGGAAAAAGTG  720
       AspSerGlyAlaSerAlaIleValIleValSerAsnPheAlaHisThrLeuGluLysVal
721  GTTGATAAAACCGCCCGTTCAGCACGTAATTCTGACCCGTATGCGCGATCAGCTATCTACG  780
       ValAspLysThrAlaValGlnHisValIleLeuThrArgMetGlyAspGlnLeuSerThr
781  GCAAAAGGCACGGTAGTCAATTTCGTTGTTAAATACATCAAGCGTTTGGTGCCGAAATAC  840
       AlaLysGlyThrValValAsnPheValValLysTyrIleLysArgLeuValProLysTyr
841  CATCTGCCAGATGCCATTTCATTTCGTAGCGCACTGCATAACGGCTACCGGATGCAGTAC  900
       HisLeuProAspAlaIleSerPheArgSerAlaLeuHisAsnGlyTyrArgMetGlnTyr
901  GTCAAACCCGAACTGGTGCCGGAAGATTTAGCTTTTCTGCAATACACCGGCGGCACCACT  960
       ValLysProGluLeuValProGluAspLeuAlaPheLeuGlnTyrThrGlyGlyThrThr
961  GGTGTGGCGAAAGGCGCGATGCTGACTCACCGCAATATGCTGGCGAACCTGGAACAGGTT 1020
       GlyValAlaLysGlyAlaMetLeuThrHisArgAsnMetLeuAlaAsnLeuGluGlnVal
1021 AACGCGACCTATGGTCCGCTGTTGCATCCGGGCAAAGAGCTGGTGGTGACGGCGCTGCCG 1080
       AsnAlaThrTyrGlyProLeuLeuHisProGlyLysGluLeuValValThrAlaLeuPro
1081 CTGTATCACATTTTTGCCCTGACCATTAACTGCCTGCTGTTTATCGAACTGGGTGGGCAG 1140
       LeuTyrHisIlePheAlaLeuThrIleAsnCysLeuLeuPheIleGluLeuGlyGlyGln
1141 AACCTGCTTATCACTAACCCGCGCGATATTCCAGGGTTGGTAAAACAGTTAGCGAAATAT 1200
       AsnLeuLeuIleThrAsnProArgAspIleProGlyLeuValLysGluLeuAlaLysTyr
1201 CCGTTTACCGCTATCACGGGCGTTAACACCTTGTTCAATGCGTTGCTGAACAATAAAGAG 1260
       ProPheThrAlaIleThrGlyValAsnThrLeuPheAsnAlaLeuLeuAsnAsnLysGlu
1261 TTCCAGCAGCTGGATTTCTCCAGTCTGCATCTTTCCGCAGGCGGTGGGATGCCAGTGCAG 1320
       PheGlnGlnLeuAspPheSerSerLeuHisLeuSerAlaGlyGlyGlyMetProValGln
1321 CAAGTGGTGCAGAGCGTTGGGTGAAACTGACCGGACAGTATCTGCTGGAAGGCTATGGC 1380
       GlnValValAlaGluArgTrpValLysLeuThrGlyGlnTyrLeuLeuGluGlyTyrGly
1381 CTTACCGAGTGTGCGCCCGCTGGTCAGCGTTAACCCATATGATATTGATTATCATAGTGGT 1440
       LeuThrGluCysAlaProLeuValSerValAsnProTyrAspIleAspTyrHisSerGly
1441 AGCATCGGTTTGCCCGTGCCGTCGACGGAAGCCAAACTGGTGGATGATGATGATAATGAA 1500
       SerIleGlyLeuProValProSerThrGluAlaLysLeuValAspAspAspAspAsnGlu
1501 GTACCACCAGGTCAACCGGGTGAGCTTTGTGTCAAAGGACCGCAGCTGATGCTGGGTTAC 1560
       ValProProGlyGlnProGlyGluLeuCysValLysGlyProGlnValMetLeuGlyTyr
1561 TGGCAGCGTCCCGATGCTACCGATGAAATCATCAAAAATGGCTGGTTACACACCGGCGAC 1620
       TrpGlnArgProAspAlaThrAspGluIleIleLysAsnGlyTrpLeuHisThrGlyAsp
1621 ATCGCCGTAATGGATGAAGAAGGATTCCTGCGCATTGTCGATCGTAAAAAAGACATGATT 1680
       IleAlaValMetAspGluGluGlyPheLeuArgIleValAspArgLysLysAspMetIle
1681 CTGGTTTCCGGTTTTAACGTCTATCCAACGAGATTGAAGATGTCGTCATGCAGCATCGTG 1740
       LeuValSerGlyPheAsnValTyrProThrArgLeuLysMetSerSerCysSerMetVal
1741 GCGTACACGAAGTCGCGGCTGTTGGCGTACCTTCCTGGCTCCAGTGGTGAAGGCGTGAAA 1800
       AlaTyrArgLysSerArgLeuLeuAlaTyrLeuProGlySerSerGlyGluGlyValLys
1801 ATCTTCGTAGTGAAAAAAGATCCATCGCTTACCGAACAGTCACTGGTGACTTTTTGCCGC 1860
       IlePheValValLysLysAspProSerLeuThrGluGluSerLeuValThrPheCysArg
1861 CGTCAGCTCACGGGATACAAAGTACCGAAGCTGGTGGAGTTTCGTGATGAGTTACCGAAA 1920
       ArgGlnLeuThrGlyTyrLysValProLysLeuValGluPheArgAspGluLeuProLys
1921 TCTAACGTCGGAAAAATTTTGCGACGAGAATTACGTGACGAAGCGCGGCAAAGTGGACAA 1980
       SerAsnValGlyLysIleLeuArgArgGluLeuArgAspGluAlaArgGlnSerGlyGln
1981 TAAAGCCTGAGCGTTAAGTCAGTCGTCAGACGCCGGTTAATCCGGCGTTTTTTTTGACGC 2040
       End
2041 CCACTAAAGAGAAAACAATTTGAATTACCAAATGATTACCACGGACGATGCGCTGGCTTC 2100
2101 TTTGTGTGAAGCCGTCCGTGCCTTTCCGGCGATAGCCCTGGATACTGAATTTGTTCGTAC 2160
2161 GCGCACTTATTACCCCGCAGCCGGGGTTGATTCAACTTTTCGATGGCGAGCATCTGGCGCT 2220
2221 AATCGGATCC 2230
```

FIG. 4. **Nucleotide sequence of the *fadD*⁺ gene and the deduced amino acid sequence for acyl-CoA synthetase.** *+1* refers to the transcriptional start defined by primer extension, potential −10 and −35 regions are noted, O_{D1} and O_{D2} are the *fadD*⁺ operators that bind FadR that are *underlined* and *italicized*, and the potential ribosome binding site sequence is noted with a *double underline*. The *underlined* amino acid sequence represents the N-terminal amino acid sequence of acyl-CoA synthetase determined by automated protein sequencing. The inverted repeat followed by 8 thymidine residues that may represent the ρ-independent terminator is between nucleotides 2010 and 2035 and is *underlined*.

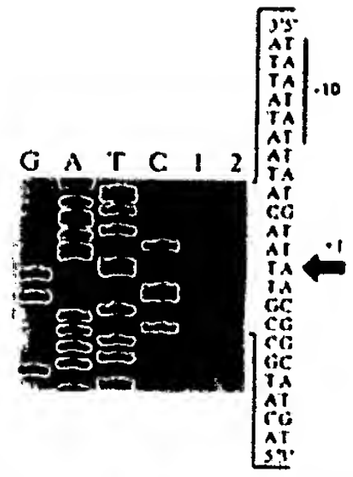

FIG. 5. **Primer extension of the 5′ end-labeled *fadD*⁺-specific oligonucleotide 5′-GATAACGGTTAAGCCAAACC-3′.** *Lanes G, A, T,* and *C* represent the *fadD* sequence using the same oligonucleotide as a primer; *lane 1* represents the primer extended fragment using RNA (20 μg) from the *fadR* strain RS3010, while *lane 2* represents the primer extended fragment using RNA (20 μg) from strain K-12. The sequence to the *right* indicates the transcriptional start at the adenine residue (noted by the *shaded arrow* and designated *+1*).

22,000 protein identified using the T7 system was shown to be contained on the *Hin*dIII-*Nco*I restriction fragment of pN300 (upstream to *fadD*⁺; data not shown). This reading frame was followed by a sequence that resembled a ρ-independent terminator. Subcloning experiments demonstrated that elimination of this fragment of DNA did not affect complementation (both by growth on oleate and acyl-CoA synthetase levels) of the *fadD88* mutation, and thus this polypeptide will not be considered further in this work.

Primer extension using the the 5′ end-labeled oligonucleotide 5′-GATAACGGTTAAGCCAAACC-3′ defined the start of transcription of the *fadD* gene at the adenine residue at nucleotide position 246 (Fig. 5). The alignment was confirmed by primer extension of a second *fadD*-specific oligonucleotide (5′-CTACCAGAGATTGATAACGG-3′)(data not shown). By convention, this residue has been designated +1 (Fig. 4). Potential −10 and −35 regions were noted around nucleotide positions 237 and 212, respectively, that were in reasonable agreement with those defined for σ-70 responsive promoters (31). As expected, the RNA isolated from the *fadR* strain RS3010 resulted in a higher signal following primer extension when compared to RNA isolated from the prototrophic strain K12 (Fig. 5). Following the translational stop of the *fadD* gene, a nearly perfect GC-rich inverted repeat is present followed by a series of 8 thymidine residues (nucleotides 2010 to 2035) which may act as a ρ-independent transcriptional terminator (Fig. 4).

Identification of the FadR Binding Sites, O_{D1} and O_{D2}—In the region just upstream from the transcriptional start site, a sequence was identified that shared homology to the operator region of the *fadB* gene defined as the FadR binding site (13). This region, from −13 to −29 (5′-*AGCTGGTATGATGAGTT*-3′), was identical in 12 of the 17 base pairs (*italicized* nucleotides noted above) with the sequence defined as the FadR operator of the *fadB* gene using DNase I footprinting (13). Particularly striking was the sequence CTGGT (−26 to −22) which was identical with that defined as the part of the FadR operator in both the *fadB* and *fadL* genes (13).[2] In order to address whether this region was the FadR operator site of the *fadD* gene, a 353-base pair *Sau*3A fragment containing this

[2] DiRusso, C. C., Metzger, A. K., and Heimert, T. L. (1993) *Mol. Micro.*, in press.

region was ligated into M13mp18, M13mp19, and pUC18 as described under "Experimental Procedures." Using DNA-protein gel retention assays, this fragment was shown to bind FadR with high affinity indicating it contained a FadR binding site (Fig. 6*A*). The apparent K_{eq} of this fragment for FadR was estimated to be 1×10^{-9} M. DNase I footprinting defined two FadR operator sites within the *fadD* promoter-containing fragment in pN330 (Fig. 6, *B* and *C*). The first FadR operator site, designated O_{D1}, covered 17 base pairs and included the sequence defined above at position −29 to −13. The second site at position −115 to −98 has been designated O_{D2}. Within O_{D2} was the sequence CTGGT which was also found in O_{D1} as well as the *fadB* and *fadL* operator sites. Overall, O_{D2} had 9 of 17 nucleotides in common with the proposed FadR consensus binding site (13).[2] The region protected by FadR from DNase I digestion at O_{D2} was identified at concentrations of FadR that were 10–50-fold higher that that used to identify O_{D1} indicating this site had a low binding affinity for FadR compared to O_R and O_{D1}.

Comparison of the E. coli Acyl-CoA Synthetase with the Rat Acyl-CoA Synthetase, Yeast Acyl-CoA Synthetase, and Firefly

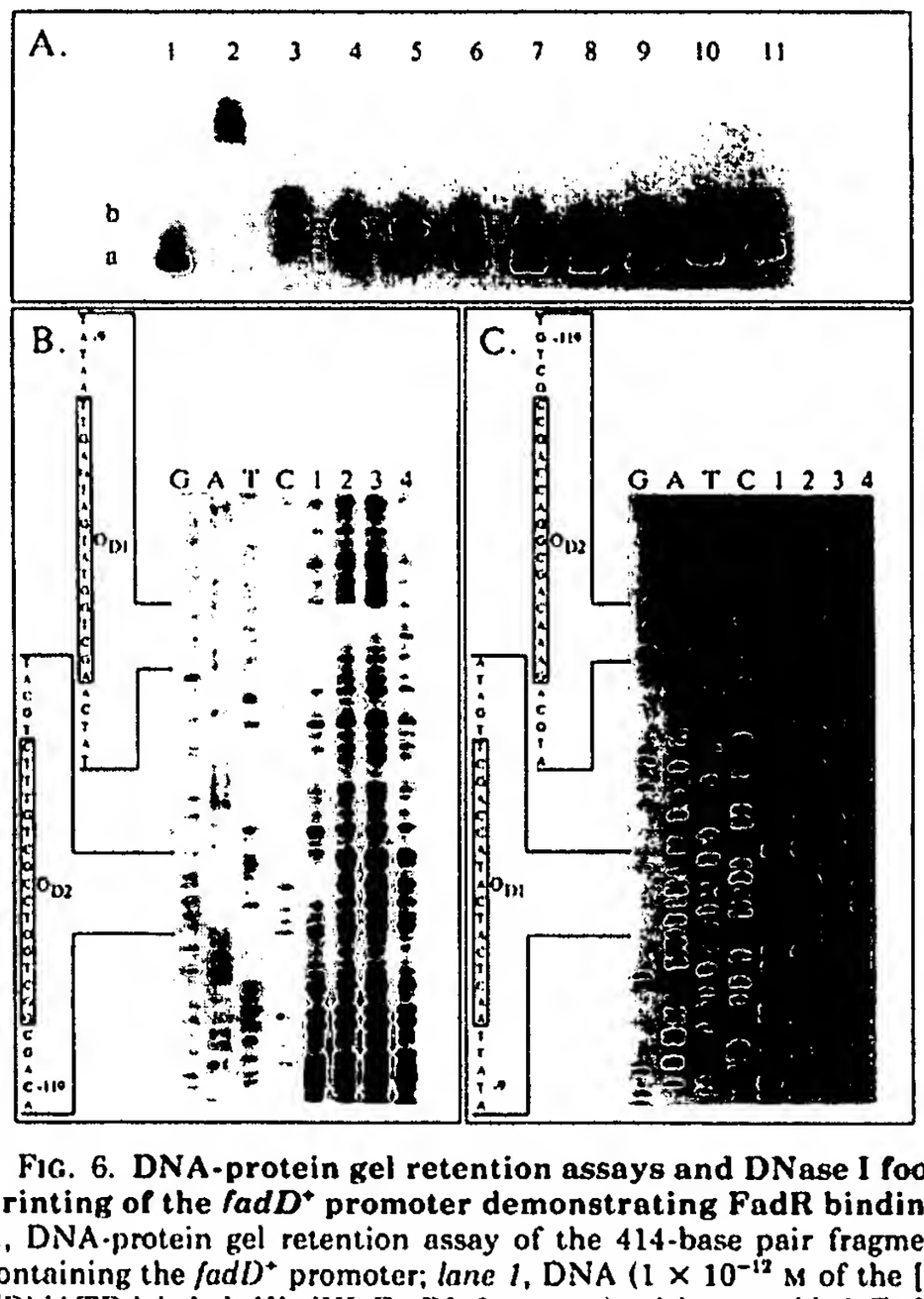

FIG. 6. **DNA-protein gel retention assays and DNase I footprinting of the *fadD*⁺ promoter demonstrating FadR binding.** *A*, DNA-protein gel retention assay of the 414-base pair fragment containing the *fadD*⁺ promoter; *lane 1*, DNA (1×10^{-12} M of the [α-^{32}P]dATP-labeled *Hin*dIII-*Eco*RI fragment) with no added FadR; *lanes 2–11* represent 1×10^{-12} M [α-^{32}P]dATP-labeled *Hin*dIII-*Eco*RI fragment and increasing concentrations of FadR; *2*, 5×10^{-8} M FadR; *3*, 1×10^{-8} M FadR; *4*, 5×10^{-9} M FadR; *5*, 1×10^{-9} M FadR; *6*, 5×10^{-10} M FadR; *7*, 1×10^{-10} M FadR; *8*, 5×10^{-11} M FadR; *9*, 1×10^{-11} M FadR; *10*, 5×10^{-12} M FadR; *11*, 1×10^{-12} M FadR; *a* refers to the DNA fragment that is not FadR-bound, and *b* represents the FadR-DNA complex; the higher molecular weight complex observed in *lane 2* is seen only at high concentrations of FadR exceeding 1×10^{-8} M. *B*, DNase I footprint of the top strand. *C*, DNase I footprint of the bottom strand. For both *B* and *C*, G, A, T, and C represent the *fadD*⁺ sequence defined using the *Eco*RI and *Hin*dIII primers, respectively, and *lanes 1*, *2*, and *3* represent decreasing concentrations of FadR (1×10^{-7} M, 2×10^{-8} M, and 4×10^{-9} M); *lane 4* represents the DNase I pattern generated in the absence of added FadR; The *fadD*⁺ operators indicated by the *open boxes* are noted as O_{D1} and O_{D2}.

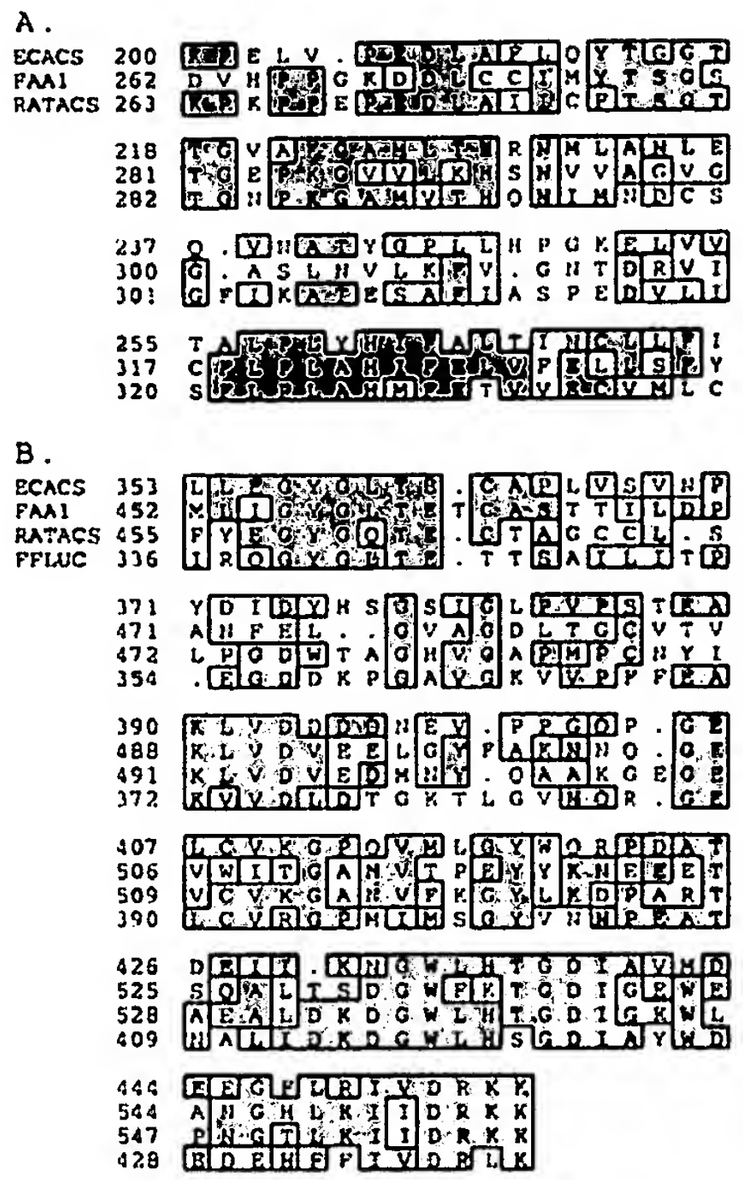

FIG. 7. **Homology between the *E. coli* acyl-CoA synthetase (*ECACS*), the rat liver acyl-CoA synthetase (*RATACS*), the yeast acyl-CoA synthetase (*FAA1*), and the firefly luciferase (*FFLC*).** *A*, similarities between ECACS, FAA1, and RATACS corresponding to amino acid residues 200–273 of the *E. coli* enzyme. *B*, similarities between ECACS, FAA1, RATACS, and FFLC corresponding to amino acid residues 353–455 of the *E. coli* enzyme. Regions with amino acid identity are *boxed* and *shaded* with *dark gray*, and those regions with high homology are *boxed* and *shaded* with *light gray*.

Luciferase—Using the Genetics Computer Group programs BESTFIT and GAP, we compared the deduced amino acid sequence from the *E. coli* acyl-CoA synthetase to that deduced for the rat (32) and the yeast enzymes (33) and firefly luciferase (34). These four enzymes were found to have extensive similarities along their entire lengths (48–51% sequence similarity when conservative amino acid substitutions are considered). Overall these four enzymes were 24–27% identical. Further analysis of these data indicated that two regions of these enzymes were highly conserved (Fig. 7). In the first region (residues 200–273 of the *E. coli* acyl-CoA synthetase), there was a 32–35% sequence identity which was extended to 53–67% similarity when conserved residues were included to residues 255–327 of the yeast enzyme and residues 262–334 of the rat enzyme. These was no apparent similarity observed between the firefly luciferase and the three acyl-CoA synthetases in this region. In the second region (amino acid residues 353–455 of the *E. coli* enzyme), 34–44% of the amino acid residues were identical and 60–65% were similar for all four enzymes (Fig. 7*B*). Suzuki *et al.* (32) proposed that for the rat enzyme, this second region may represent the ATP binding site. Part of our current research efforts are being directed at addressing whether or not this region of the *E. coli* acyl-CoA synthetase actually represents a nucleotide binding region.

DISCUSSION

In the present paper, we report the cloning, sequencing, and expression of the *fadD* gene of *E. coli* encoding acyl coenzyme

A synthetase. The *fadD* gene was identified in clone 6D1 from the Kohara gene library and subsequently shown to be encoded within a 2.2-kb *Nco*I-*Cla*I fragment of genomic DNA by complementation analysis. The expression of the *fadD* gene was monitored both by following acyl-CoA synthetase activities in the collection of *fadD* and *fadD*$^+$ plasmids and by following induction of the *fadD*$^+$ gene using T7 RNA polymerase. Acyl-CoA synthetase levels were only 2-fold inducible in the presence of the long-chain fatty acid oleate which differed from the levels of induction observed for other *fad* gene products (acyl-CoA dehydrogenase, enoyl-CoA hydratase, β-hydroxyacyl-CoA dehydrogenase, and β-ketothiolase) (5, 26). The DNA sequence of the *fadD* gene predicted a protein with 558 amino acid residues and a molecular weight of 62,028 starting with UUG as the translational initiation codon. This alignment was confirmed by N-terminal amino acid sequence analysis of purified acyl-CoA synthetase. No evidence was obtained which indicated the acyl-CoA synthetase was post-translationally processed (*i.e.* the presence of a signal sequence and/or N-terminal amino acid modification). The transcriptional initiation site of the *fadD* gene was determined to be an adenine residue 60 nucleotides upstream from the initiation site of translation using primer extension of two different *fadD*-specific oligonucleotides. The T7 RNA polymerase experiments estimated the size of the *E. coli* acyl-CoA synthetase to be 62,000 which was in agreement with that deduced from the DNA sequence and that defined from earlier reports.

The *fadD* gene contained two operator sites for the binding of FadR. The first (O_{D1}) was slightly upstream (−29 to −13) from the transcriptional start and had a relatively high affinity for FadR ($K_{eq} \sim 1 \times 10^{-9}$ M). The estimated affinity of the *fadD* operator O_{D1} toward FadR ($\sim 1 \times 10^{-9}$ M) was nearly an order of magnitude lower than that defined for the *fadB* promoter (3×10^{-10} M) (13). This operator site was appropriately positioned to block transcription when filled by FadR as it overlapped the presumptive −10 region. The second operator site (O_{D2}) was found 114 base pairs upstream from the transcriptional start (−115 to −99). This site had considerably less affinity toward FadR ($K_{eq} \geq 1 \times 10^{-8}$ M) as estimated using DNase I footprinting. Studies are being conducted to determine the precise contribution of this site as well as the contribution of O_{D1} in the expression of the *fadD* gene.

DiRusso and her colleagues (13) demonstrated that the long-chain fatty acyl-CoA molecule is the inducer of the fatty acid degradative genes by showing that inclusion of these compounds (in nanomolar concentrations) prevented FadR binding to the *fadB* operator site in DNA-protein gel retention assays while long-chain fatty acids did not (13). Due to the affinity of O_{D1}, it is expected that when this site is filled, transcription of *fadD*$^+$ is likely to be turned off or maintained at a low basal level. The role of O_{D2} is less clear; perhaps this second site regulates a second promoter upstream from the primary promoter identified here. Alternatively, there may be cooperative interaction between proteins bound at O_{D1} and O_{D2} that contribute to enhanced repression of the *fadD* promoter. When fatty acids are present in the growth media, the acyl-CoA synthetase enzymatically produces an increased intracellular pool of long-chain fatty acyl-CoA molecules that results in the derepression of transcription. The net result is coordinate induction of transcription of the genes involved in fatty acid transport, activation, and degradation, including *fadD*.

The presence of a UUG translation initiation codon for acyl-CoA synthetase was noted in the course of the present study. This initiation codon is relatively rare (found in 1% of *E. coli* genes) and in some cases acts to down-regulate the expression of a given protein (30). It is plausible that the production of acyl-CoA synthetase may also be down-regulated. In the case of the *rnd* gene of *E. coli* (encoding RNase D), replacement of the native UUG with AUG results in an 11-fold increase in RNase D expression (34). We are presently investigating whether the acyl-CoA synthetase activity is subject to comparable regulation.

Acyl-CoA synthetase is crucial for the uptake of exogenous long-chain fatty acids that are destined to be utilized as a source of carbon and metabolic energy. This enzyme has been proposed to vectorially transport long-chain fatty acids across the inner membrane with a concomitant thioesterification to the CoA derivatives. Acyl-CoA synthetase functions by generating a fatty acid-adenylate intermediate which in turn is converted into a fatty acyl-CoA. In this respect, this enzyme is likely to bind ATP (see "Discussion" below) and thus may be functionally analogous to the ATPase component of bacterial permeases that represents class of transport proteins collectively referred to as "traffic ATPases" (36). There is evidence that suggests a H^+/long-chain fatty acid co-transporter is present in the inner membrane (6, 7). If this is the case, this postulated component must interact directly with the acyl-CoA synthetase in the vectorial transport of long-chain fatty acids.

In the course of our analysis of the *fadD* gene, we compared the deduced amino acid sequence of the *E. coli* acyl-CoA synthetase to the rat liver acyl-CoA synthetase, the yeast acyl-CoA synthetase, and firefly luciferase. Although the four proteins shared significant similarity along their entire lengths, a higher degree of similarity in the amino acid sequences from these four proteins was identified toward their carboxyl ends (amino acid residues 353–455 of the *E. coli* enzyme). Due to the common mechanism of action of these proteins, it seems likely that this region may be of functional importance. Suzuki *et al.* (32) proposed that this region of the rat acyl-CoA synthetase represented the ATP binding domain. This proposal is, in part, based on the similar enzyme mechanism proposed for firefly luciferase. The firefly luciferase reaction, like the acyl-CoA synthetase reaction, proceeds by a two-step mechanism that results in the formation of an adenylated intermediate. In both the firefly luciferase and the acyl-CoA synthetase enzyme mechanisms, there is a reaction between the carboxyl group of the substrate (lucifern and long-chain fatty acid, respectively) and ATP to form the adenylated intermediate. In both cases, this activation reaction requires the pyrophosphorolysis of ATP. The formation of an adenylated intermediate requires that ATP bind transiently to the enzyme as part of the catalytic cycle (8, 32). The second region of the *E. coli* acyl-CoA synthetase that shared similarity with the yeast and rat enzymes was more centrally located along the linear amino acid sequence of these proteins (amino acid residues 200–273 of the *E. coli* enzyme). As this homology was not seen with the firefly luciferase, this region may specify a component unique to acyl-CoA synthetases (*e.g.* fatty acid binding domain and/or the coenzyme A binding domain). The significance of these similarities with respect to the function of the *E. coli* acyl-CoA synthetase is presently under investigation.

Acknowledgments—We thank Jeff Gordon, Robert Duronio, and Laura Knoll for data on the yeast acyl-CoA synthetase prior to publication, William Lane of the Harvard University Microchemistry facility for determination of the N-terminal amino acid sequence of acyl-CoA synthetase, Charles Liarakos for discussions concerning translational regulation, and William Holden for technical assistance.

REFERENCES

1. Black, P. N., Said, B., Ghoan, C., Beach, J. V., and Nunn, W. D. (1987) *J. Biol. Chem.* **262**, 1412–1419
2. Black, P. N. (1990) *Biochim. Biophys. Acta* **1046**, 97–105
3. Black, P. N. (1991) *J. Bacteriol.* **173**, 535–542
4. Kumar, G. B., and Black, P. N. (1991) *J. Biol. Chem.* **266**, 1348–1353
5. Overath, P., Pauli, G., and Schairer, H. U. (1969) *Eur. J. Biochem.* **7**, 559–574
6. Kameda, K. (1986) *Biochem. Int.* **13**, 343–350
7. Kameda, K., Suzuki, L. K., and Imai, Y. (1987) *Biochem. Int.* **14**, 227–234
8. Groot, P. H. E., Scholte, H. R., and Hulsmann, W. C., (1976) *Adv. Lipid Res.* **14**, 75–126
9. Kameda, K., and Nunn, W. D. (1981) *J. Biol. Chem.* **256**, 5702–5707
10. Kameda, K., Suzuki, L. K., and Imai, Y. (1985) *Biochim. Biophys. Acta* **840**, 29–36
11. Klein, K., Steinberg, R., Fiethen, B., and Overath, P. (1971) *Eur. J. Biochem.* **19**, 442–450
12. DiRusso, C. C. (1988) *Nucleic Acids Res.* **16**, 7995–8009
13. DiRusso, C. C., Heimert, T. L., and Metzger, A. K. (1992) *J. Biol. Chem.* **267**, 8685–8691
14. Simons, R. W., Egan, P. A., Chute, H. T., and Nunn, W. D. (1980) *J. Bacteriol.* **142**, 621–632
15. Nunn, W. D., Colburn, R., and Black, P. N. (1986) *J. Biol. Chem.* **261**, 167–171
16. Berman, M. L., Enquist, L. W., and Silhavy, T. J. (1981) *Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
17. Miller, J. H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
18. Kohara, Y., Akiyama, K., and Isono, K. (1987) *Cell* **50**, 495–508
19. Black, P. N., Kianian, S. K., DiRusso, C. C., and Nunn, W. D. (1985) *J. Biol. Chem.* **260**, 1780–1789
20. Del Sal, G., Manfioletti, G., and Schneider, C. (1989) *BioTechniques* **7**, 514–519
21. Chang, A. Y. C., and Cohen, S. N. (1978) *J. Bacteriol.* **134**, 1141–1156
22. Maniatis, T., Friesch, E. F., and Sambrook, J. (1981) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
23. Ish-Horowitz, D., and Burke, J. F. (1981) *Nucleic Acids Res.* **9**, 2989–2998
24. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
25. Bradford, M. (1976) *Anal. Biochem.* **72**, 248–254
26. DiRusso, C. C. (1990) *J. Bacteriol.* **172**, 6459–6468
27. Studier, W. F., and Moffat, B. A. (1986) *J. Mol. Biol.* **189**, 113–130
28. Laemmli, U. K. (1970) *Nature* **227**, 680–685
29. Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035–10038
30. Gold, L. (1988) *Annu. Rev. Biochem.* **57**, 199–233
31. Hawley, D. K., and McClure, W. R. (1983) *Nucleic Acids Res.* **11**, 2237–2255
32. Suzuki, H., Kawarabayasi, Y., Kondo, J., Abe, T., Nishikawa, K., Kimura, S. Hashimoto, T., and Yamamoto, T. (1990) *J. Biol. Chem.* **265**, 8681–8685
33. Duronio, R. J., Knoll, L. J., and Gordon, J. I. (1992) *J. Cell Biol.* **117**, 531–538
34. deWet, J. R., Wood, K. V., DeLuca, M., Helenski, D. R., and Subramani, S. (1987) *Mol. Cell. Biol.* **7**, 725–737
35. Zhang, J., and Deutscher, M. P. (1989) *J. Biol. Chem.* **264**, 18228–18233
36. Ames, G. F.-L. (1990) in *The Bacteria: Bacterial Energetics* (Krulwich, T. A., ed) Vol. 12, pp. 225–246, Academic Press, New York

van Beilen

Molecular Microbiology (1992) 6(21), 3121–3136

DNA sequence determination and functional characterization of the OCT-plasmid-encoded *alkJKL* genes of *Pseudomonas oleovorans*

Jan B. van Beilen,[1] Gerrit Eggink,[2] Hans Enequist,[1] Rolf Bos[1] and Bernard Witholt[1]*
[1]*Department of Biochemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands.*
[2]*ATO Agrotechnology, Postbus 17, 6700 AA Wageningen, The Netherlands.*

Summary

The *alkBFGHJKL* and *alkST* operons encode enzymes that allow *Pseudomonas putida (oleovorans)* to metabolize alkanes. In this paper we report the nucleotide sequence of a 4592 bp region of the *alkBFGHJKL* operon encoding the AlkJ, AlkK and AlkL polypeptides.

The *alkJ* gene encodes a protein of 59 kilodaltons. The predicted amino acid sequence shows significant homology with four flavin proteins: choline dehydrogenase, a glucose dehydrogenase and two oxidases. AlkJ is membrane-bound and converts aliphatic medium-chain-length alcohols into aldehydes. The properties of AlkJ suggest that it is linked to the electron transfer chain. AlkJ is necessary for growth on alkanes only in *P. putida* alcohol dehydrogenase (AlcA) mutants.

AlkK is homologous to a range of proteins which act by an ATP-dependent covalent binding of AMP to their substrate. This list includes the acetate, coumarate and long-chain fatty acid CoA ligases. The *alkK* gene complements a *fadD* mutation in *Escherichia coli*, which shows that it indeed encodes an acyl-CoA synthetase. AlkK is a 60 kilodalton protein located in the cytoplasm.

AlkL is homologous to OmpW, a *Vibrio cholerae* outer membrane protein of unknown function, and a hypothetical polypeptide encoded by *ytt4* in *E. coli*. AlkL, OmpW and Ytt4 all have a signal peptide and end with a sequence characteristic of outer membrane proteins. The *alkL* gene product was found in the outer membrane of *E. coli* W3110 containing the *alk*-genes. The *alkL* gene can be deleted without a clear effect on growth rate. Its function remains unknown.

The G+C content of the *alkJKL* genes is 45%, identical to that of the *alkBFGH* genes, and significantly lower than the G+C content of the OCT-plasmid and the *P. putida* chromosome.

Received 15 May, 1992; revised and accepted 9 July, 1992. *Present address for correspondence: Biotechnology Institute, Swiss Federal Institute of Technology Zürich, Hönggerberg HPT, CH-8093 Zürich, Switzerland. For correspondence. Tel. (1) 377 2088; Fax (1) 371 0658.

Introduction

The OCT-plasmid (Chakrabarty *et al.*, 1973) allows *Pseudomonas oleovorans* to utilize alkanes as sole carbon and energy source. Two operons located on this plasmid; *alkBFGHJKL* (Eggink *et al.*, 1987a; Kok *et al.*, 1989b) and *alkST* (Eggink *et al.*, 1988; 1990) encode the enzymes necessary to convert alkanes to fatty acids.

The first step in alkane catabolism; hydroxylation of alkanes to alkanols, is catalysed by the alkane hydroxylase system (Peterson *et al.*, 1966). This enzyme system consists of the *alkB*, *alkG* and *alkT* gene products alkane hydroxylase, rubredoxin and rubredoxin reductase (Kok *et al.*, 1989a,b; Eggink *et al.*, 1990). The *alkH* gene encodes an NAD-dependent aldehyde dehydrogenase (Kok *et al.*, 1989b), and AlkS regulates expression of the *alkBFGHJKL* operon (Eggink *et al.*, 1988). AlkF is a non-functional rubredoxin (Kok *et al.*, 1989b).

The distal part of the operon, earlier known as *alkC* (Owen *et al.*, 1984), encodes three peptides identified by Eggink *et al.* (1987a), and designated AlkJ (58 kDa), AlkK (59 kDa) and AlkL (20 kDa) (Kok *et al.*, 1989b). Marker rescue experiments of alcohol dehydrogenase mutants (Owen *et al.*, 1984) mapped the alcohol dehydrogenase function to the *alkJ* gene. The alcohol dehydrogenase activity was lost in pGEc41 where the distal three cistrons of the *alkBFGHJKL* operon were deleted (Eggink *et al.*, 1987b). No mutations that affect alkane utilization have been mapped to the *alkK* and *alkL* genes. The above data are summarized in Fig. 1A and B.

In this paper we present the nucleotide sequence of the *alkJKL* region and provide evidence for the role of the peptides encoded by this region in alkane oxidation.

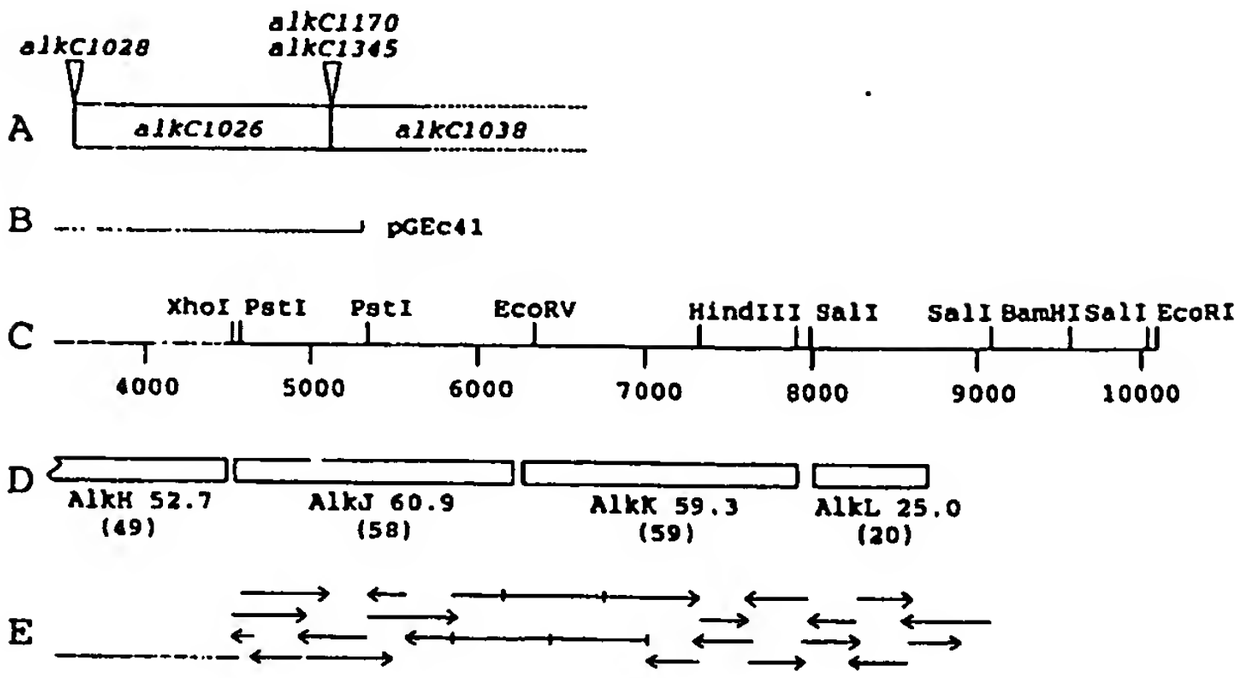

Fig. 1. Genetic structure of the *alkJKL* region and strategy in determining its sequence.
A. Previous genetic experiments indicate that *alkJ* encodes an NAD-independent dye-linked alcohol dehydrogenase. Triangles indicate Tn*7* insertions which inactivate the alcohol dehydrogenase (Fennewald *et al.*, 1979), and boxes indicate the approximate positions of two-point mutations in the alcohol dehydrogenase gene (as determined by Owen *et al.*, 1984 and relative to the DNA sequence published by Kok *et al.*, 1989a,b).
B. Location of 3' end of the insert of pGEc41.
C. Position of relevant restriction sites. Previous R-loop experiments indicated that the mRNA ends around position 8100, plus or minus 600 bp (Eggink *et al.*, 1987a).
D. Location of the *alkJ*, *alkK*, *alkL* and the neighbouring *alkH* open reading frames, the inferred molecular masses ($\times 10^{-3}$) of AlkH, AlkJ, AlkK and AlkL, and (in parenthesis) the apparent molecular masses ($\times 10^{-3}$) of these peptides, as determined in minicell expression experiments (Eggink *et al.*, 1987a).
E. Sequence strategy. Part of the sequence was determined by primer walking (vertical bars). The arrows indicate the direction and extent of sequence determination of individual clones.

Results

Nucleotide sequence of the alkJKL *region*

The nucleotide sequence of the *alkJKL* region was determined following the strategy depicted in Fig. 1E. Three open reading frames (ORFs) were found (Fig. 2). The position and length of these ORFs, the predicted sizes of the corresponding encoded peptides, potential ribosomal binding sites and G+C content are shown in Table 1. No other large ORFs could be detected downstream of the AlkL ORF in a sequence of 377 nucleotides. The sequences of the *alkBFGHJKL* operon and its translation products AlkJ, AlkK and AlkL have been submitted to the EMBL Data Library, and are available under accession number X65936 (*P. oleovorans* genomic DNA + OCT plasmid).

Comparison of sequence information to previous R-loop and peptide mapping experiments

We have previously analysed the *alkBFGHJKL* operon for translation products in *E. coli* minicells (Eggink *et al.*,

Table I. Properties of the *alkJ*, *alkK* and *alkL* genes and gene products.

	alkJ	*alkK*	*alkL*
Sequence			
Coding region	4548–6221	6284–7921	8026–8715
Size of protein	558 residues (60.9 kDa)	546 residues (59.3 kDa)	230 (203) residues 25.0 (22.1) kDa
G+C content of region	45.6%	45.4%	43.0%
Ribosome-binding site	CGAGAA(6)AUG	UGAGG(7)AUG	CGAGGG(7)AUG
Distance to preceding open-reading-frame	41 bp	63 bp	104 bp
Minicell experiments[a]			
Size of protein[b]	58 kDa	59 kDa	20 kDa
Mapped between[b]	4500–6150	6150–8000 (*Sal*I)	7921 (*Hin*dIII)-9092
Genetic and biochemical data			
Size of protein	59 kDa	60 kDa	25 kDa
alk mutations[c]	*alkC1026, alkC1038*	None	None
Function	Alcohol dehydrogenase	Acyl-CoA synthetase	Unknown
Localization	Cytoplasmic membrane (peripheral)	Cytoplasm	Outer membrane

a. Eggink *et al.* (1987a).
b. based on co-ordinates and restriction sites in Eggink *et al.* (1987a).
c. Owen *et al.* (1984).

```
                4506                                                    4548
GTT AAG TAT TTG AGC   TGAGTAATTTTTTCGATAAATCATTACCTCGAGAAGATAAA ATG TAC GAC TAT  455
Val Lys Tyr Leu Ser                                               MET Tyr Asp Tyr  4

ATA ATC GTT GGT GCT GGA TCT GCA GGA TGT GTG CTT GCT AAT CGT CTT TCG GCC GAC CCC  461
Ile Ile Val Gly Ala Gly Ser Ala Gly Cys Val Leu Ala Asn Arg Leu Ser Ala Asp Pro  24

TCT AAA AGA GTT TGT TTA CTT GAA GCT GGG CCG CGA GAT ACG AAT CCG CTA ATT CAT ATG  467
Ser Lys Arg Val Cys Leu Leu Glu Ala Gly Pro Arg Asp Thr Asn Pro Leu Ile His Met  44

CCG TTA GGT ATT GCT TTG CTT TCA AAT AGT AAA AAG TTG AAT TGG GCT TTT CAA ACT GCG  473
Pro Leu Gly Ile Ala Leu Leu Ser Asn Ser Lys Lys Leu Asn Trp Ala Phe Gln Thr Ala  64

CCA CAG CAA AAT CTC AAC GGC CGG AGC CTT TTC TGG CCA CGA GGA AAA ACG TTA GGT GGT  479
Pro Gln Gln Asn Leu Asn Gly Arg Ser Leu Phe Trp Pro Arg Gly Lys Thr Leu Gly Gly  84

TCA AGC TCA ATC AAC GCA ATG GTC TAT ATC CGA GGG CAT GAA GAC GAT TAC CAC GCA TGG  485
Ser Ser Ser Ile Asn Ala Met Val Tyr Ile Arg Gly His Glu Asp Asp Tyr His Ala Trp  104

GAG CAG GCG GCC GGC CGC TAC TGG GGT TGG TAC CGG GCT CTT GAG TTG TTC AAA AGG CTT  491
Glu Gln Ala Ala Gly Arg Tyr Trp Gly Trp Tyr Arg Ala Leu Glu Leu Phe Lys Arg Leu  124

GAA TGC AAC CAG CGA TTC GAT AAG TCC GAG CAC CAT GGG GTT GAC GGA GAA TTA GCT GTT  497
Glu Cys Asn Gln Arg Phe Asp Lys Ser Glu His His Gly Val Asp Gly Glu Leu Ala Val  144

AGT GAT TTA AAA TAT ATC AAT CCG CTT AGC AAA GCA TTC GTG CAA GCC GGC ATG GAG GCC  503
Ser Asp Leu Lys Tyr Ile Asn Pro Leu Ser Lys Ala Phe Val Gln Ala Gly Met Glu Ala  164

AAT ATT AAT TTC AAC GGA GAT TTC AAC GGC GAG TAC CAG GAC GGC GTA GGG TTC TAT CAA  509
Asn Ile Asn Phe Asn Gly Asp Phe Asn Gly Glu Tyr Gln Asp Gly Val Gly Phe Tyr Gln  184

GTA ACC CAA AAA AAT GGA CAA CGC TGG AGC TCG GCG CGT GCA TTC TTG CAC GGT GTA CTT  515
Val Thr Gln Lys Asn Gly Gln Arg Trp Ser Ser Ala Arg Ala Phe Leu His Gly Val Leu  204

TCC AGA CCA AAT CTA GAC ATC ATT ACT GAT GCG CAT GCA TCA AAA ATT CTT TTT GAA GAC  521
Ser Arg Pro Asn Leu Asp Ile Ile Thr Asp Ala His Ala Ser Lys Ile Leu Phe Glu Asp  224

CGT AAG GCG GTT GGT GTT TCT TAT ATA AAG AAA AAT ATG CAC CAT CAA GTC AAG ACA ACG  527
Arg Lys Ala Val Gly Val Ser Tyr Ile Lys Lys Asn Met His His Gln Val Lys Thr Thr  244

AGT GGT GGT GAA GTA CTT CTT AGT CTT GGC GCA GTC GGC ACG CCT CAC CTT CTA ATG CTT  533
Ser Gly Gly Glu Val Leu Leu Ser Leu Gly Ala Val Gly Thr Pro His Leu Leu Met Leu  264

TCT GGT GTT GGG GCT GCA GCC GAG CTT AAG GAA CAT GGT GTT TCT CTA GTC CAT GAT CTT  539
Ser Gly Val Gly Ala Ala Ala Glu Leu Lys Glu His Gly Val Ser Leu Val His Asp Leu  284

CCT GAG GTG GGG AAA AAT CTT CAA GAT CAT TTG GAC ATC ACA TTG ATG TGC GCA GCA AAT  545
Pro Glu Val Gly Lys Asn Leu Gln Asp His Leu Asp Ile Thr Leu Met Cys Ala Ala Asn  304

TCG AGA GAG CCG ATA GGT GTT GCT CTT TCT TTC ATC CCT CGT GGT GTC TCG GGT TTG TTT  551
Ser Arg Glu Pro Ile Gly Val Ala Leu Ser Phe Ile Pro Arg Gly Val Ser Gly Leu Phe  324

TCA TAT GTG TTT AAG CGC GAG GGG TTT CTC ACT AGT AAC GTG GCA GAG TCG GGT GGT TTT  557
Ser Tyr Val Phe Lys Arg Glu Gly Phe Leu Thr Ser Asn Val Ala Glu Ser Gly Gly Phe  344

GTA AAA AGT TCT CCT GAT CGT GAT CGG CCC AAT TTG CAG TTT CAT TTC CTT CCA ACT TAT  563
Val Lys Ser Ser Pro Asp Arg Asp Arg Pro Asn Leu Gln Phe His Phe Leu Pro Thr Tyr  364

CTT AAA GAT CAC GGT CGA AAA ATA GCG GGT GGT TAT GGT TAT ACG CTA CAT ATA TGT GAT  569
Leu Lys Asp His Gly Arg Lys Ile Ala Gly Gly Tyr Gly Tyr Thr Leu His Ile Cys Asp  384

CTT TTG CCT AAG AGC CGA GGC AGA ATT GGC CTA AAA AGC GCC AAT CCA TTA CAG CCG CCT  575
Leu Leu Pro Lys Ser Arg Gly Arg Ile Gly Leu Lys Ser Ala Asn Pro Leu Gln Pro Pro  404

TTA ATT GAC CCG AAC TAT CTT AGC GAT CAT GAA GAT ATT AAA ACC ATG ATT GCG GGT ATT  581
Leu Ile Asp Pro Asn Tyr Leu Ser Asp His Glu Asp Ile Lys Thr Met Ile Ala Gly Ile  424

AAG ATA GGG CGC GCT ATT TTG CAG GCC CCA TCG ATG GCG AAG CAT TTT AAG CAT GAA GTA  587
Lys Ile Gly Arg Ala Ile Leu Gln Ala Pro Ser Met Ala Lys His Phe Lys His Glu Val  444

GTA CCG GGC CAG GCT GTT AAA ACT GAT GAT GAA ATA ATC GAA GAT ATT CGT AGG CGA GCT  593
Val Pro Gly Gln Ala Val Lys Thr Asp Asp Glu Ile Ile Glu Asp Ile Arg Arg Arg Ala  464

GAG ACT ATA TAC CAT CCG GTA GGT ACT TGT AGC ATG GGT AAA GAT CCA GCG TCA GTT GTT  599
Glu Thr Ile Tyr His Pro Val Gly Thr Cys Arg Met Gly Lys Asp Pro Ala Ser Val Val  484

GAT CCG TGC CTG AAG ATC CGT GGG TTG GCA AAT ATT AGA GTC GTT GAT GCG TCA ATT ATG  605
Asp Pro Cys Leu Lys Ile Arg Gly Leu Ala Asn Ile Arg Val Val Asp Ala Ser Ile Met  504
```

Fig. 2.

```
CCG CAC TTG GTC GCG GGT AAC ACA AAC GCT CCA ACT ATT ATG ATT GCA GAA AAT GCG GCA  611
Pro His Leu Val Ala Gly Asn Thr Asn Ala Pro Thr Ile Met Ile Ala Glu Asn Ala Ala  524

GAA ATA ATT ATG CGG AAT CTT GAT GTG GAA GCA TTA GAG GCT AGC GCT GAG TTT GCT CGC  617
Glu Ile Ile Met Arg Asn Leu Asp Val Glu Ala Leu Glu Ala Ser Ala Glu Phe Ala Arg  544
                                                  6221
GAG GGT GCA GAG CTA GAG TTG GCC ATG ATA GCT GTC TGC ATG TAAAAAACATGGTCAATAGATGG  624
Glu Gly Ala Glu Leu Glu Leu Ala Met Ile Ala Val Cys Met  558
                                          6284
TTTTTTAATGAACATAAATCATCAATGTGAGGCGACGTG ATG TTA GGT CAG ATG ATG CGT AAT CAG TTG  631
                                        MET Leu Gly Gln Met Met Arg Asn Gln Leu  10

GTC ATT GGT TCG CTT GTT GAG CAT GCT GCA CGA TAT CAT GGT GCG AGA GAG GTG GTT TCA  637
Val Ile Gly Ser Leu Val Glu His Ala Ala Arg Tyr His Gly Ala Arg Glu Val Val Ser  30

GTC GAA ACC TCT GGA GAA GTA ACA AGA AGT TGT TGG AAA GAA GTG GAG CTT CGT GCT CGT  643
Val Glu Thr Ser Gly Glu Val Thr Arg Ser Cys Trp Lys Glu Val Glu Leu Arg Ala Arg  50

AAG CTC GCT TCT GCA TTG GGC AAG ATG GGT CTT ACG CCT AGT GAT CGT TGT GCA ACG ATT  649
Lys Leu Ala Ser Ala Leu Gly Lys Met Gly Leu Thr Pro Ser Asp Arg Cys Ala Thr Ile  70

GCA TGG AAC AAT ATT CGT CAT CTT GAG GTT TAC TAC GCT GTC TCT GGC GCA GGA ATG GTA  655
Ala Trp Asn Asn Ile Arg His Leu Glu Val Tyr Tyr Ala Val Ser Gly Ala Gly Met Val  90

TGC CAT ACA ATC AAT CCG AGG CTT TTC ATT GAG CAG ATC ACA TAT GTG ATA AAC CAT GCG  661
Cys His Thr Ile Asn Pro Arg Leu Phe Ile Glu Gln Ile Thr Tyr Val Ile Asn His Ala  110

GAG GAT AAG GTA GTA CTT CTT GAT GAT ACG TTC TTG CCA ATC ATT GCT GAG ATT CAC GGT  667
Glu Asp Lys Val Val Leu Leu Asp Asp Thr Phe Leu Pro Ile Ile Ala Glu Ile His Gly  130

TCG TTA CCA AAA GTC AAG GCG TTT GTC TTG ATG GCT CAT AAT AAT TCA AAT GCA TCT GCT  763
Ser Leu Pro Lys Val Lys Ala Phe Val Leu Met Ala His Asn Asn Ser Asn Ala Ser Ala  150

CAA ATG CCA GGA TTG ATT GCA TAC GAG GAT CTA ATT GGT CAG GGT GAT GAT AAC TAT ATA  679
Gln Met Pro Gly Leu Ile Ala Tyr Glu Asp Leu Ile Gly Gln Gly Asp Asp Asn Tyr Ile  170

TGG CCT GAT GTA GAT GAA AAT GAG GCG TCT AGT CTA TGT TAC ACA TCA GGT ACT ACG GGC  685
Trp Pro Asp Val Asp Glu Asn Glu Ala Ser Ser Leu Cys Tyr Thr Ser Gly Thr Thr Gly  190

AAC CCG AAG GGT GTA CTT TAT TCA CAC CGC TCG ACA GTT TTG CAT TCA ATG ACC ACC GCA  691
Asn Pro Lys Gly Val Leu Tyr Ser His Arg Ser Thr Val Leu His Ser Met Thr Thr Ala  210

ATG CCA GAC ACA CTA AAT TTG TCT GCG CGA GAT ACC ATT TTG CCC GTA GTT CCA ATG TTT  697
Met Pro Asp Thr Leu Asn Leu Ser Ala Arg Asp Thr Ile Leu Pro Val Val Pro Met Phe  230

CAT GTA AAT GCA TGG GGG ACT CCA TAT TCC GCT GCA ATG GTT GGT GCG AAG CTA GTT CTT  703
His Val Asn Ala Trp Gly Thr Pro Tyr Ser Ala Ala Met Val Gly Ala Lys Leu Val Leu  250

CCT GGT CCG GCT CTT GAT GGC GCT AGT TTA TCG AAG TTG ATT GCT AGC GAA GGA GTT AGC  709
Pro Gly Pro Ala Leu Asp Gly Ala Ser Leu Ser Lys Leu Ile Ala Ser Glu Gly Val Ser  270

ATT GCT CTT GGG GTG CCG GTT GTT TGG CAG GGG TTG TTA GCG GCA CAA GCC GGT AAT GGT  715
Ile Ala Leu Gly Val Pro Val Val Trp Gln Gly Leu Leu Ala Ala Gln Ala Gly Asn Gly  290

TCT AAA AGC CAA AGC CTC ACG CGG GTT GTT GTA GGA GGT TCG GCC TGT CCT GCG TCT ATG  721
Ser Lys Ser Gln Ser Leu Thr Arg Val Val Val Gly Gly Ser Ala Cys Pro Ala Ser Met  310

ATT AGA GAA TTT AAC GAT ATA TAT GGT GTT GAA GTT ATT CAT GCT TGG GGT ATG ACT GAG  727
Ile Arg Glu Phe Asn Asp Ile Tyr Gly Val Glu Val Ile His Ala Trp Gly Met Thr Glu  330

CTT TCG CCA TTT GGC ACG GCA AAC ACT CCA CTC GCG CAC CAC GTA GAT TTA TCT CCA GAT  733
Leu Ser Pro Phe Gly Thr Ala Asn Thr Pro Leu Ala His His Val Asp Leu Ser Pro Asp  350

GAA AAG CTT TCA CTG CGC AAA AGC CAA GGG CGC CCG CCT TAC GGT GTC GAG TTA AAA ATC  739
Glu Lys Leu Ser Leu Arg Lys Ser Gln Gly Arg Pro Pro Tyr Gly Val Glu Leu Lys Ile  370

GTT AAT GAT GAG GGG ATT AGA CTA CCT GAA GAT GGT CGA AGT AAA GGC AAC CTA ATG GCG  745
Val Asn Asp Glu Gly Ile Arg Leu Pro Glu Asp Gly Arg Ser Lys Gly Asn Leu Met Ala  390

CGT GGG CAC TGG GTT ATT AAA GAT TAC TTT CAT AGC GAT CCT GGT TCG ACA CTC TCA GAT  751
Arg Gly His Trp Val Ile Lys Asp Tyr Phe His Ser Asp Pro Gly Ser Thr Leu Ser Asp  410

GGT TGG TTT TCA ACT GGA GAC GTG GCT ACC ATA GAT TCG GAC GGT TTC ATG ACA ATC TGT  757
Gly Trp Phe Ser Thr Gly Asp Val Ala Thr Ile Asp Ser Asp Gly Phe Met Thr Ile Cys  430

GAT CGT GCA AAG GAC ATT ATA AAG TCT GGC GGT GAG TGG ATC AGT ACG GTA GAG CTG GAG  763
Asp Arg Ala Lys Asp Ile Ile Lys Ser Gly Gly Glu Trp Ile Ser Thr Val Glu Leu Glu  450
```

Fig. 2.

```
AGT ATT GCG ATT GCG CAC CCT CAT ATT GTT GAT GCT GCT GTT ATA GCT GCA AGG CAC GAA  769
Ser Ile Ala Ile Ala His Pro His Ile Val Asp Ala Ala Val Ile Ala Ala Arg His Glu  470

AAA TGG GAC GAG CGA CCT CTC CTC ATC GCA GTT AAA TCC CCT AAT TCG GAA TTA ACA AGT  775
Lys Trp Asp Glu Arg Pro Leu Leu Ile Ala Val Lys Ser Pro Asn Ser Glu Leu Thr Ser  490

GGT GAG GTA TGT AAT TAT TTC GCA GAT AAG GTG GCT AGA TGG CAA ATT CCA GAT GCC GCT  781
Gly Glu Val Cys Asn Tyr Phe Ala Asp Lys Val Ala Arg Trp Gln Ile Pro Asp Ala Ala  510

ATC TTT GTT GAA GAA CTG CCA CGC AAT GGT ACT GGC AAG ATT TTG AAG AAT CGT TTG CGC  787
Ile Phe Val Glu Glu Leu Pro Arg Asn Gly Thr Gly Lys Ile Leu Lys Asn Arg Leu Arg  530
                                                        7921
GAG AAA TAT GGT GAT ATT TTA TTG CGC AGT AGT TCT TCT GTC TGT GAA TAA GCTTTCTGTAT  793
Glu Lys Tyr Gly Asp Ile Leu Leu Arg Ser Ser Ser Ser Val Cys Glu  546

GGGCTTTGACTGATTTTTTAGATATCAGTCTCTGTGACATGTTAGCAGTTCTATTTAAGAATATGTCGACAACAAAACG  801
          8026
AGGGTAGCACA ATG AGT TTT TCT AAT TAT AAA GTA ATC GCG ATG CCG GTG TTG GTT GCT AAT  807
            MET Ser Phe Ser Asn Tyr Lys Val Ile Ala Met Pro Val Leu Val Ala Asn  17

TTT GTT TTG GGG GCG GCC ACT GCA TGG GCG AAT GAA AAT TAT CCG GCG AAA TCT GCT GGC  813
Phe Val Leu Gly Ala Ala Thr Ala Trp Ala Asn Glu Asn Tyr Pro Ala Lys Ser Ala Gly  37

TAT AAT CAG GGT GAC TGG GTC GCT AGC TTC AAT TTT TCT AAG GTC TAT GTG GGT GAG GAG  819
Tyr Asn Gln Gly Asp Trp Val Ala Ser Phe Asn Phe Ser Lys Val Tyr Val Gly Glu Glu  57

CTT GGC GAT CTA AAT GTT GGA GGG GGG GCT TTG CCA AAT GCT GAT GTA AGT ATT GGT AAT  825
Leu Gly Asp Leu Asn Val Gly Gly Gly Ala Leu Pro Asn Ala Asp Val Ser Ile Gly Asn  77

GAT ACA ACA CTT ACG TTT GAT ATC GCC TAT TTT GTT AGC TCA AAT ATA GCG GTG GAT TTT  831
Asp Thr Thr Leu Thr Phe Asp Ile Ala Tyr Phe Val Ser Ser Asn Ile Ala Val Asp Phe  97

TTT GTT GGG GTG CCA GCT AGG GCT AAA TTT CAA GGT GAG AAA TCA ATC TCC TCG CTG GGA  837
Phe Val Gly Val Pro Ala Arg Ala Lys Phe Gln Gly Glu Lys Ser Ile Ser Ser Leu Gly  117

AGA GTC AGT GAA GTT GAT TAC GGC CCT GCA ATT CTT TCG CTT CAA TAT CAT TAC GAT AGC  843
Arg Val Ser Glu Val Asp Tyr Gly Pro Ala Ile Leu Ser Leu Gln Tyr His Tyr Asp Ser  137

TTT GAG CGA CTT TAT CCA TAT GTT GGG GTT GGT GTT GGT CGG GTG CTA TTT TTT GAT AAA  849
Phe Glu Arg Leu Tyr Pro Tyr Val Gly Val Gly Val Gly Arg Val Leu Phe Phe Asp Lys  157

ACC GAC GGT GCT TTG AGT TCG TTT GAT ATT AAG GAT AAA TGG GCG CCT GCT TTT CAG GTT  855
Thr Asp Gly Ala Leu Ser Ser Phe Asp Ile Lys Asp Lys Trp Ala Pro Ala Phe Gln Val  177

GGC CTT AGA TAT GAC CTT GGT AAC TCA TGG ATG CTA AAT TCA GAT GTG CGT TAT ATT CCT  861
Gly Leu Arg Tyr Asp Leu Gly Asn Ser Trp Met Leu Asn Ser Asp Val Arg Tyr Ile Pro  197

TTC AAA ACG GAC GTC ACA GGT ACT CTT GGC CCG GTT CCT GTT TCT ACT AAA ATT GAG GTT  867
Phe Lys Thr Asp Val Thr Gly Thr Leu Gly Pro Val Pro Val Ser Thr Lys Ile Glu Val  217
                                                    8715
GAT CCT TTC ATT CTC AGT CTT GGT GCG TCA TAT GTT TTC TAA GTAATCAGGTCTGTCACTGTCGC  874
Asp Pro Phe Ile Leu Ser Leu Gly Ala Ser Tyr Val Phe  230

AGCACTCTTTCGATCCATTAGTGTGCCGTCTCACAAATAATGCTCCTCATCCCTGTGAGACCATTGCCTCACGGGCTCG  882

CTGTATCTTGTTCAAAATATCTTCCCCCTTGGCGTTCCAGACGTAGCGAGTCGGCTGCGCATTTCGCAGTGCCAGGAAC  889

GTGGTAATCGAGCTTTCCAGTTCGCGAACCGAACTGAAACTGCCATCACGCAGGTACACCGTGATATCGCGAAAGAAGC  897

GTTCAACCATGTTCATCCACGAACTGGAGGTCGGGGTGAAATGCATGTGGAAGCGTTTGTGCTTCTCAAGCCACGCCCT  905

CACTTTGGGATGCTTGTGCGTGGCGTAGTTGTCGA  9092
```

Fig. 2. Nucleotide sequence of the *alkJKL* region. This sequence continues that reported in Kok *et al.* (1989a,b). Nucleotide positions of the first and last nucleotide of the ORFs are shown above the sequence. Nucleotide and amino acid positions are shown to the right of the sequences. The putative Shine and Dalgarno boxes are underlined (Shine and Dalgarno, 1975). These sequence data appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Data Libraries under the accession number X65936.

1987a). Starting from the promotor the detected polypeptides had molecular masses of 41 (AlkB), 15 (AlkF), 49 (AlkH), 58 (AlkJ), 59 (AlkK), and 20 kDa (AlkL), respectively. The positions of the ORFs and the calculated molecular masses of the peptides encoded by the *alkJKL* ORFs are shown in Fig. 1D. Both gene location and inferred peptide size are in accordance with the minicell experiments, except for a difference of 5 kDa in the molecular mass of AlkL, as can be seen in Table 1.

According to R-loop experiments the *alkBFGHJKL* operon ends at position 8100 ± 600 bp (Eggink *et al.*, 1987a). The most distal position (8700) is close to the end of AlkL ORF at position 8715.

In our earlier minicell experiments the *alkJ* cistron encoded two peptides with molecular masses of 58 kDa or 37 kDa (Eggink *et al.*, 1987a). The 37 kDa peptide

```
                                                                                          10        20        30        40        50
AlkJ                                                                                  MYDYIIVGAGSAGCVLANRLS--ADPSKRVCLLEAGPRD--TNPLIHMPLGIAL
BetA                                                                                  MQFDYIIIGAGSAGNVLATRLT--EDPNTSVLLLEAGGPDYRFDFRTQMPAALAF
GLD   MSASASACDCLVGVPTGPTLASTCGGSAFMLFMGLLEVFIRSQCDLEDPCGRASSRFRSEPDYEYDFIVIGGGSAGSVVASRLS--EVPQWKVLLIEAGGDEPVGAQIPSMFLNFIG
AOX                                                                                 MAIPDEFDIIVVGGGSTGCCIAGRLANLDDQNLTVALIEGGENNINNPWVYLPGVYPRN
GOD                                     MQTLLVSSLVVSLAAALPHYIRSNGIEASLLTDPKDVSGRTVDYIIAGGGLTGLTTAARLT--ENPNISVLVIESGSYESDRGPIIEDLNAYGD
                                                                                      * *. *.* .*  * **  . .    * ..*.* .

           60        70        80        90       100       110       120       130                140       150
AlkJ  LSNSKKLNWAFQTAPQQNLNGRSLFWPRGKTLGGSSSINAMVYIRGHEDDYHAWEQAAG-RYWGWYRALELFKRLECNQRFDK----------SEHHGVDGELAVSDLK---YINPL
BetA  PLQGKRYNWAYETEPEPFMNNRRMECGRGKGLGGSSLINGMCYIRGNALDLDNWAQEPGLENWSYLDCLPYYRKAETRDMGE----------NDYHGGDGPVSVTTSK--PGVNPL
GLD   SDIDYRYNTEPEPMACLSSMEQRCYWPRGKVLGGTSVLNGMMYVRGNREDYDDWAAD-GNPGWAYNDVLPFFKKSEDNLDLDE--------VGTEYHAKGGLLPVGKF---PYNPPL
AOX   MRLDSKTATFYSSRPSKALNGRRAIVPCANILGGGSSINFLMYTRASASDYDDWES----EGWSTDELLPLIKKIETYQRPCN---------NRDLHGFDGPIKVSFGN---YTYPT
GOD   IFGSSVDHAYETVEL--ATNNQTALIRSGNGLGGSTLVNGGTWTRPHKAQVDSWETVFGNEGWNWDNVAAYSLQAERARAPNAKQIAAGHYFNASCHGVNGTVHAGPRDTGDDYSPI
                      .. ***.. .*   ..*   . ..*.        *.  .            *                     *. .* . ..               *

           160       170        180       190       200        210       220           230       240             250
AlkJ  SKAFVQAGMEANINFNGDFNGEY--QDGVGFYQVTQK-NGQRWSSARAFLHG-VLSRPNLDIITDAHASKILFED----RKAVGVSYIKKNMHHQVKTT-----SGGEVLLSLGAVG
BetA  FEAMIEAGVQAGYPRTDDLNGYQ--QEGFGPMDRTVTPQGRRASTARGYLDQ-AKSRPNLTIRTHAMTDHIIFDG----KRAVGVEWLEGDSTIPTRAT-----ANKEVLLCAGAIA
GLD   SYAILKAGEELGFSVH-DLNGQN--STGFMIAQMTAR-NGIRYSSARAFLRP-ARMRNNLHILLNTTATKILIHP--HTKNVLGVEV-SDQFGSTRKIL-----VKKEVVLSAGAVN
AOX   CQDFLRAAESQGIPVVDDLEDFKTSHGAEHWLKWINRDLGRRSDSAHAYVHPTMRNKQSLFLITSTKCDKVIIED----GKAVAVRTVPMKPLNPKKPVSRTFRARKQIVISCGTIS
GOD   VKALMSAVEDRGVPTKKDFGCGD--PHGVSMFPNTLHEDQVRSDAAREWLLP-NYQRPNLQVLTGQYVGKVLLSQNGTTPRAVGVVEFGTHKGNTHNVY-----AKHEVLLAAGSAV
         *     .     *..     .       *      * ..*....     . .* .     .....     ....*     .                 .... *.

         260       270       280       290       300       310       320          330       340       350
AlkJ  TPHLLMLSGVGAAAELKEHGVSLVHDLPEVGKNLQDHLDITLMCAANSREPIGVALSFIPRGVSGLFS--YVFKREGFLTSNVAESGGFVKSSPDRD-----------------RPN
BetA  SPQILQRSGVGNAELLAEFDIPLVHELPGVGENLQDHLEMYLQYECKEPVSLYPALQWWNQPKIGAE---WLFGGTGVGASNHFEAGGFIRSREE-----------------FAWPN
GLD   SPHILLLSGVGPKDELQQVNVRTVHNLPGVGKNLHNHVTYFTNFFIDDADTAPLNWATAMEYLLFRD---GLMSGTGISDVTAKLATRYADSPE--------------------RPD
AOX   SPLVLQRSGIGAAHHLRSVGVKPIVDLPGVGENFQDHYCFFTPYYVKPDVPTFDDFVRGDPVAQKAAFDQWYSNKDGPLTTNGIEAGVKIRPTEEELATADEDFRRGYAEYFENKPD
GOD   SPTILEYSGIGMKSILEPLGIDTVVDLP-VGLNLQDQTTATVRSRITSAGAGQGQAAWFATFNETF----GDYSEKAHELLNTKLEQWAEEAVARGGFHNTTALLI----QYENYRD
      .* .*  **.*    *    ..  . .** ** *....                                 .       .        .                        ..

        360       370              380       390       400       410       420       430       440       450
AlkJ  LQFHFLPTYLKDHGRKIAGG------YGYTLHICDLLPKSRGRIGLKSANP-LQPPLIDPNYLSDHEDIKTMIAGIKIGRAILQAPSMAKHFKHEVVPGQAVKT------------
BetA  IQYHFLPVAINYNGSNAVKE------HGFQCHVGSMRSPSRGHVRIKSRDP-HQHPAILFNYMSHEQDWQEFRDAIRITREIMHQPALDQYRGREISPGVECQT------------
GLD   LQLYFGGYLASCARTGQVGELLSNNSRSIQIFPAVLNPRSRGFIGLRSADP-LEPPRIVANYLTHERDVKTLVEGIKFVIRLSQTTPLKQYGMRLDKTVVKGCEAHAFGS-------
AOX   KPLMHYSVISGFFGDHTKIPN----GKFMTMFHFLEYPFSRGFVRITSANP-YDAPDFDPGFLNDERDLWPMVWAYKKSRETARRMES---FAGEVTSHHPLFKVDSPARARDLDLE
GOD   WIVNHNVAYSELFLDTAGV---------ASFDVWDLLPFTRGYVHILDKDPYLHHFAYDPQYFLNELDLLGQAAATQLARNISNSGAMQTYFAGETIPGDNLAY------------
        .                            . .** . . . .*     .  .    .  .*     . .     .

                                                          460       470       480              490       500       510
AlkJ  ------------------------------------------------DDEIIEDIRRRAETIYHPVGTCRMGKDPAS-------VVDPCLKIRGLANIRVVDASIMPHLVAGN
BetA  ------------------------------------------------DEQLDEFVRNHAETAFHPCGTCKMGYDEMS-------VVDGEGRVHGLEGLRVVDASIMPQIITGN
GLD   ------------------------------------------------DAYWECAVRQNTGPENHQAGSCKMGPSHDP-----MAVVNHELRVHGIRGLRVMDTSIMPKVSSGN
AOX   TCSAYAGPKHLTANLYHGSWTVDIDKPTPKNDFHVTSNQVQLHDIEYTEEDDEAIVNYIKEHTETTWHCLGTCSMAPREGSKIAPKGGVLDARLNVYGVQNLKVADLSVCPDNVGCN
GOD   ---------------------------------------------DADLSAWTEYIPYHFRPNYHGVGTCSMMPKEM------GGVVDNAARVYGVQGLRVIDGSIPPTQMSSH
                                                  .    .    .    .    *   *.* *            *..    .. *.   .** *.*. *    ..

          520       530       540       550
AlkJ  TNAPTIMIAENAAEIIMRNLDVEALEASAEFAREGAELELAMIAVCM 558
BetA  LNATTIMIGEKIADMI-RGQBALPRSTAGYFVANGMPVRAKK      556
GLD   THAPAVMIAEKGAYLLKRAWGAKV                        612
AOX   TYSTALTIGEKAATLVAEDLGYSGSDLDMTIPNFRLGTYEETGLARF 664
GOD   VMTVFYAMALKISDAILEDYASMQ                        605
      . .   .. . . . .
```

% amino acid identity	and length of homologous region
100	(558)
37.3	(550)
33.3	(546)
27.2	(536)
22.6	(536)

Fig. 3. Comparison of AlkJ with choline dehydrogenase, glucose dehydrogenase, alcohol oxidase and glucose oxidase. The amino acid sequences of *E. coli* choline dehydrogenase (BetA), *Drosophila melanogaster* glucose dehydrogenase (GLD), *Hansenula polymorpha* yeast alcohol oxidase (AOX), and *Asper-gillus nidulans* glucose oxidase (GOD) were aligned using the programs FASTP (Lipman and Pearson, 1985) and CLUSTAL, which is part of the PC/GENE package. The shaded region shows the location of the nucleotide binding βαβ loops in these proteins (Fig. 5). Also shown are the percentage of identity of BetA, GLD, AOX, and GOD, with AlkJ, and the length of the region that could be aligned directly using FASTP.

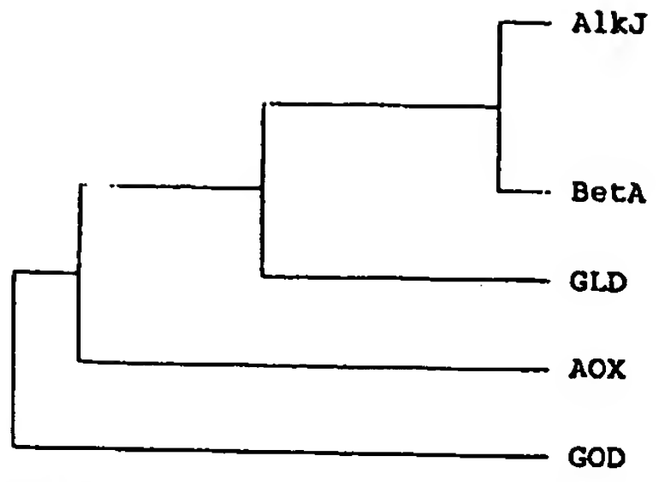

Fig. 4. Dendrogram of the alignment of AlkJ with dehydrogenases and oxidases. The distance to the branching points is a measure for relative evolutionary distance. The figure was constructed using the PC/GENE program CLUSTAL.

was clearly derived from the 58 kDa peptide, as shown by limited proteolysis with *Staphylococcus aureus* V8 protease. Subsequent experiments have shown that in a number of plasmids, all derived from pGEc59 and pGEc60, a deletion of one base (A) from a stretch of As had occurred, resulting in an ORF of 371 amino acids (40.7 kDa). Plasmids pGEc50 and pGEc51 were not derived from pGEc59 or pGEc60 and produced a 58 kDa protein in minicell experiments (Eggink *et al.*, 1987a). The sequence of Fig. 2 shows that this region encodes a polypeptide with a calculated molecular mass of 60.9 kDa.

Nucleotide composition and codon usage of the alkJKL *genes*

The average G+C content of the *alkJKL* genes is 45 %, which is comparable with that of the *alkBFGH* genes (46–47 %) (Kok *et al.*, 1989b), but much lower than the 62–67 % found for the *P. putida* chromosome (Mandel, 1966) and the IncP2 plasmids including OCT (Fennewald *et al.*, 1978). The region following the *alkL* stop codon shows an increase in G+C content to 54 %, clearly higher than the G+C content of the *alk* genes. Together with the absence of an ORF downstream of AlkL, and the R-loop experiments (Eggink *et al.*, 1987a), this indicates that *alkL* is indeed the last gene of the *alkBFGHJKL* operon.

The unusual nucleotide composition is reflected in the codon usage of these genes. Unlike most *Pseudomonas* genes the *alk*-genes show no preference for G+C over A+T at the wobble position (data not shown). As this does not prevent AlkB from being produced to high levels in *Pseudomonas* and *E. coli*, the unusual codon usage apparently does not affect translation significantly.

Primary structure of AlkJ

We compared the primary structure of AlkJ with the protein sequences compiled in the Protein Sequence Data Base (Dayhoff *et al.*, 1983; Lipman and Pearson, 1985). AlkJ is homologous with *E. coli* choline dehydrogenase (BetA; Lamark *et al.*, 1991), *Drosophila melanogaster* and *D. pseudoobscura* glucose dehydrogenase (GLD; Krasney *et al.*, 1990), *Hansenula polymorpha* alcohol oxidase (AOX; Ledeboer *et al.*, 1985) and *Aspergillus niger* glucose oxidase (GOD; Frederick *et al.*, 1990). An alignment of these sequences, and the homology to AlkJ expressed as a percentage positional sequence identity, is shown in Fig. 3. The relative evolutionary distance between these proteins is shown as a dendrogram in Fig. 4.

All of these proteins, including AlkJ, possess a characteristic fingerprint for ADP binding (Wierenga *et al.*, 1986) at, or close to, their amino terminus (Fig. 5). The two oxidases and glucose dehydrogenase are flavin proteins. In case of BetA this has not yet been determined. However, since these proteins are all NAD(P)-independent and possess the same fingerprint it is likely that BetA and AlkJ are flavin proteins as well.

The ADP-binding fingerprint consists of 11 amino acid positions at which specific amino acids occur (Fig. 5). Sequences that contain the expected amino acids at only nine or 10 of the positions can still be ADP-binding βαβ-folds. Interestingly, AlkJ, BetA, GLD, AOX and GOD all have an aspartic acid residue at the first position of the fingerprint, a deviation from the consensus. In fact, we found that more than half of the flavin proteins in the Swiss-Prot database (Release 18) that have a recognizable ADP-binding fold (about 60), have an aspartate at this position. At the second position of the fingerprint, AlkJ, BetA and GOD have a tyrosine residue. This again is not in agreement with the consensus fingerprint, but it is also found in proteins such as the glutathione and mercury reductases.

ADP binding fingerprint		●	■	-	■	-	G	-	G	-	-	G	-	-	-	■	-	-	■		(loop)			■	-	■	-	o
AlkJ	3	D	Y	I	V	V	G	A	G	S	A	G	C	V	L	A	N	R	L	20	(7)	28	V	C	L	L	E	32
BetA	4	D	Y	I	I	I	G	A	G	S	A	G	N	V	L	A	T	R	L	21	(7)	29	V	L	L	L	E	33
GLD	66	D	F	I	V	I	G	G	G	S	A	G	S	V	V	A	S	R	L	82	(7)	90	V	L	L	I	E	95
AOX	8	D	E	I	V	V	G	G	G	S	T	G	C	C	I	A	G	R	L	25	(9)	35	V	A	L	I	E	39
GOD	43	D	Y	I	I	A	G	G	G	L	T	G	L	T	G	A	A	R	L	60	(7)	68	V	L	V	I	E	72

Fig. 5. Potential ADP-binding fold in AlkJ, BetA, GLD, AOX and GOD. The consensus sequence that can fold in a βαβ-structure with ADP-binding properties (Wierenga *et al.*, 1986) is shown in the top row. The symbols indicate which amino acids should occur at the given position: (●) K, R, H, S, T, Q, N; (■) A, I, L, V, M, C, G; (O) D, E. The loop represents an amino acid stretch with variable length (between parenthesis). Shading indicates agreement with the consensus sequence. The binding sites are located at or near the amino terminus.

amino acids 185–193 of AlkK	T	S	G	T	T	G	N	P	K
putative AMP-binding pattern	T S G	S T G	G	T S	T S	G	X	P Z	K	Z=LIVM
P-loop in ras-proteins			G	X	X	G A	X	G	K	
P-loop in adenylate cyclases			G	X	P	G	X	G	K	G T

Fig. 6. Comparison of amino acids 185–193 in AlkK and the conserved sequence in adenylate-forming enzymes (the putative AMP-binding pattern), with P-loop sequences in two protein families. The shaded residues are conserved between most of the sequences. X is any amino acid. The underlined G is conserved between all different P-loops (Saraste *et al.*, 1990), but is in most cases replaced by a proline in the putative AMP-binding pattern (Bairoch, 1992).

Phenotype of strains equipped with recombinant alk *plasmids; function of AlkJ*

The OCT-plasmid-encoded particulate alcohol dehydrogenase AlcO was studied by Benson and Shapiro (1976) and localized to the *alkC* region of the *alkBAC* operon by marker-rescue experiments (Owen *et al.*, 1984). The mutations and transposon insertions that affect the expression of the alcohol dehydrogenase are shown in Fig. 1A. In minicell experiments the alcohol dehydrogenase function appeared to correspond to a 58 kDa polypeptide encoded by *alkC* (Eggink *et al.*, 1987a), which was later named AlkJ (Kok *et al.*, 1989b). To confirm this assignment we deleted a large part of the *alkJ* ORF. *P. putida* and *E. coli* strains containing recombinant *alk* plasmids were plated on minimal medium and tested for growth on octane or 1-octanol vapour. Table 2 shows the results obtained for strains carrying pGEc47ΔJ, wild-type *P. oleovorans* (GPo1) and recombinant strains described earlier in Eggink *et al.* (1987b). Like pGEc41 (pGEc47ΔJKL), pGEc47ΔJ is unable to complement the alcohol dehydrogenase mutations of *P. putida* PpS81 and PpS597, while pGEc47 allows these mutants to grow on *n*-octane. This shows that *alkJ* encodes the alcohol dehydrogenase function.

The growth on octane of *E. coli* GEc137 carrying pGEc47ΔJ is similar to that of *E. coli* GEc137 carrying pGEc47, and somewhat faster than that of *E. coli* GEc137 carrying pGEc41. This indicates that in *E. coli* AlkJ is not essential for growth on alkanes.

Alcohol dehydrogenase activity of recombinant strains

Since AlkJ showed homology to two oxidases, we tested the oxidase activity of *P. putida* GPo12 and *E. coli* GEc137, carrying pGEc47 and pGEc47ΔJ, using a peroxidase/2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid (ABTS) assay (Verduyn *et al.*, 1984). However, none of these *E. coli* or *P. putida* recombinants showed production of H_2O_2, even after prolonged incubations. The presence of 1 mM Na-azide to inhibit catalase activity did not affect the outcome of the test.

In previous studies Benson and Shapiro (1976) found that AlcO (now AlkJ) activity is NAD(P) independent, is located in the pellet after centrifugation at 48000 × g, and that it could be assayed using the artificial electron acceptor phenazine methosulphate (PMS). It has these properties in common with choline dehydrogenase (BetA) of *E. coli*, which shows 37% identity with AlkJ. BetA is strictly O_2-dependent, but it does not produce H_2O_2. When Triton X-100 is added the O_2-dependent activity of BetA is destroyed. However, the solubilized enzyme still catalyses the oxidation of its substrate in the presence of PMS. This has led the authors to propose that BetA is electron transfer-linked (Lamark *et al.*, 1991). We therefore carried out experiments to determine whether AlkJ is electron transfer-dependent as well.

Total membranes were isolated from several of the strains described above. The membrane preparations were assayed for PMS/2,6-dichlorophenol-indophenol (DCPIP)-dependent activities, and for O_2-dependent alcohol dehydrogenase activity using a Biological Oxygen Monitor. Under these conditions the half-life of AlkJ activity is less than 30 min, and unlike BetA, AlkJ is completely inactive with PMS when Triton X-100 is added to a concentration of 1%. The AlkJ activity was ≈ 0.4 – 1.0 U mg^{-1} protein in several experiments, whether PMS/DCPIP oxidation or O_2 consumption was measured. In the absence of the *alkJ* gene, oxidation levels were lowered to ≈ 10 –15 %, similar to the effect of adding X-100. These results show that AlkJ does not transfer electrons from the substrate to a soluble cofactor, but to oxygen. Since no H_2O_2 is produced, oxygen must be reduced to water, most likely through the electron transfer chain.

Primary structure of AlkK

Comparison of AlkK with the EMBL databank sequences revealed that AlkK shows clear amino acid similarity with a number of proteins that activate their substrate by covalently binding AMP to a carboxy acid group of the substrate. Turgay *et al.* (1992) have constructed a phylogenetic tree for this family of proteins which shows three major branches (a, b, c). Branch a contains two acetyl-CoA synthetases; branch b contains, among others, two 4-coumarate CoA ligases (25 % sequence identity), long-chain fatty acid CoA ligase, firefly luciferase, and EntE (an

enzyme which activates 2,3-dihydroxybenzoate to the corresponding acyladenylate); and branch c contains 22 domains of proteins that activate amino acids or amino acid analogues, and are involved in the synthesis of secondary metabolites such as siderophores and antibiotics. AlkK is most closely related to the 4-coumarate CoA ligases in branch b (25 % sequence identity). The similarity of AlkK to the other proteins or protein domains ranges from 17 to 24% identity.

When all proteins or protein domains are aligned, very few residues are perfectly conserved. Only one small region is relatively well conserved in all proteins, including AlkK (amino acids 182–193), as was also found by Scholten *et al.* (1991). The sequence pattern was found simultaneously by Bairoch (1992), who noted that it is similar to ATP-binding P-loops from different protein families (Saraste *et al.*, 1990), even though an invariant glycine is replaced by a proline (Fig. 6). As all proteins that are related to AlkK adenylate their substrates using ATP, this sequence motif may be involved in binding of ATP or AMP. Bairoch (1992), has added it to the PROSITE dictionary of sites and patterns found in protein sequences as PROSITE PS00455, which is a putative AMP-binding domain signature.

Function of AlkK

The *alkBFGHJ/alkT* genes have been shown to specify a complete pathway from alkane to fatty acid. Since *P. putida* is able to grow on fatty acids we have assumed previously that subsequent oxidation steps, starting with the activation of fatty acids to the corresponding acyl-CoAs, were encoded by the chromosome. However, the fact that AlkK is related to a number of acyl-CoA synthetases suggested that it might have a similar function.

We tested whether the *alkK* gene complemented an acyl-CoA synthetase (*fadD*) mutation in *E. coli*. In this mutant the other *fad* genes can no longer be induced, because the inducer, which is either a long-chain acyl-CoA, or long-chain fatty acid, is not formed or transported into the cytosol. If AlkK is able to activate or transport long-chain fatty acids it should complement the *fadD* mutation completely. However, if AlkK is specific for medium-chain-length fatty acids, growth would not occur because the *fad* genes are not induced. In this case a *fadR* mutation, which causes constitutive expression of the *fad* genes, is necessary to obtain growth. Fortunately, these mutants appear very frequently when *E. coli* is plated on minimal media with decanoate as a sole carbon source (Overath *et al.*, 1969), or when *E. coli* strains carrying pGEc47 are plated on minimal media with octane as the sole carbon and energy source (Eggink *et al.*, 1987b).

E. coli K27 (*fadD*) was transformed with plasmids pGEc47, pGEc41, pGEc47ΔK and pGEc47ΔJ. Plasmids pGEc47 and pGEc47ΔJ contain the *alkK* gene, whereas pGEc41 and pGEc47ΔK do not. Plasmid pGEc47ΔJ was included to test whether deletion of a part of *alkJ* still allows expression of *alkK*. The recombinants were plated on minimal medium, and incubated in octane vapour, or plated on minimal medium containing 10 mM Na-oleate, and incubated in the presence of dicyclopropylketone vapour to induce the *alk* system. Initially, very weak growth was observed with *E. coli* K27 (pGEc47) and K27 (pGEc47ΔJ) on both types of media. No growth was observed with K27, K27 (pGEc47ΔK) and K27 (pGEc41). After about 7 d a limited number of large colonies appeared on the plates with K27 (pGEc47) and K27 (pGEc47ΔJ) incubated in octane vapour. These colonies probably represented *fadR* mutants. The colonies were cured of their plasmid and restreaked on oleate-containing plates to show that they did not represent revertants of the *fadD* mutation: no growth was observed. The cured strain K27 *fadR*, named GEc354, was then retransformed with all four plasmids, and plated on minimal media. GEc354 (pGEc47) grew well when incubated in octane vapour for 3 d. Growth of GEc354 (pGEc47ΔJ) was somewhat slower. No growth was observed for GEc354 (pGEc41) and GEc354 (pGEc47ΔK) (Table 2).

The results indicate that AlkK is an acyl-CoA synthetase, with a specificity for medium-chain-length fatty acids. This specificity is not unexpected since the other *alk*-system enzymes are specific for medium-chain-length substrates as well.

Primary structure of AlkL

AlkL shows 28% identity in a 170-amino-acid overlap with the *Vibrio cholerae* outer membrane protein OmpW (Jalajakumari and Manning, 1990), and Ytt4 (Stoltzfus *et al.*, 1988), a hypothetical 22.9 kDa protein in the *trpA–tonB* intergenic region of *E. coli* (Fig. 7). The first 60 amino acids of AlkL show no homology with OmpW or Ytt4. The amino-terminal sequences of the three proteins strongly resemble bacterial signal sequences, while the remainder of the proteins is hydrophilic. AlkL has three closely spaced potential signal peptidase cleavage sites, of which the third, between amino acids 27 and 28, scores highest by the rules proposed by Von Heyne (1986). The calculated molecular mass of AlkL after cleavage of the putative signal peptide at this position is 22138 daltons, which is reasonably close to the apparent molecular mass of 20 kDa found in the minicell experiments.

All three proteins end with a pattern characteristic of outer membrane proteins (Struyvé *et al.*, 1991). This pattern has characteristics suggesting that the carboxy-terminal 10 residues form a membrane-spanning amphipathic β-sheet. When the three sequences were analysed for turn-prone regions according to Paul and Rosenbusch (1985), it appeared that many of these regions are

```
AlkL   MSFSNYKVIAMPVLVANFVLGAATAWANENYPAKSAGYNQGDWVASFNFS   50
OmpW   M-----KQ-TICLAVLAALLAAPV------FAHQEGDF----------IV   28
Ytt4   M-----KKLTVAALAVTTLLSGSA------FAHEAGEF----------FM   29
       *     *  ..   . . .*....      .. .....           .

          1               2               3               4
AlkL   KVYVGEELGDLNVGGGALPNADVSIGNDTTLTFDIAYFVSSNIAVDFFVG  100
OmpW   RAGIASVVPNDSSDKVLNTQSELAVNSNTHLGLTLGYMFTDNISFEVLAR   78
Ytt4   RAGSATVRPTEGAGGTLGSLGGFSVTNNTQLGLTFTYMATDNIGVELLAA   79
       ..  ..   . . .   . .. .....* *.....*. ..**. . ..

                        5               6               7
AlkL   VPARAKFQGEKS-ISSLGRVSEVDYGPAILSLQYHYDSFERLY-PYVGVG  148
OmpW   TPFSHKISTSGGELGSLGDIGETKHLPPTFMVQYYFGEANSTNRPYVGAG  128
Ytt4   TPFRHKIGTRAT-----GDIATVHHLPPTLMAQWYFGDASSKFRPYVGAG  124
       .* . *. . .     * ..... *...  *..... ..   ****.*

                        8               9              10
AlkL   VGRVLFFDK------TDGALSSFDIKDKWAPAFQVGLRYDLGNSWMLNSD  192
OmpW   LNYTTFFDESFNSTGTNNALSDLKLDDSWGLAANVGFDYMLNDSWFLNAY  178
Ytt4   INYTTFFDNGFNDHGKEAGLSDLSLKDSWGAAGQVGVDYLINRDWLVNMS  174
       .. . ***.      ....**.....*.*. * .**  * .. .*..*

                  11              12
AlkL   VRYIPFKTDVTGTLGPVPVSTKIEVDPFILSLGASYVF   230
OmpW   VWYANIETTATYKAGADAKSTDVEINPWVFIIAGGYKF   216
Ytt4   VWYMDIDTTANYKLGGAQQHDSVRLDPWVFMFSAGYRF   212
       *.*  ..*... . *  .  ... ..*... ....* *

Identity  : 38 (16 %)
Similarity: 107 (45 %)
```

Fig. 7. Comparison of AlkL with OmpW and Ytt4. The amino acid sequence of AlkL was aligned with the sequence of the *V. cholerae* outer membrane protein OmpW, and with the hypothetical *E. coli* protein Ytt4. Identical residues are indicated by (*), conserved residues are indicated by (.). Shaded regions represent turns predicted with the algorithm of Paul and Rosenbusch (1985). The numbers indicate putative transmembrane β-sheet strands. The YXF motif characteristic of the carboxy terminus of the outer membrance proteins is indicated in bold.

located at corresponding positions in AlkL, OmpW and Ytt4 (Fig. 7, shaded residues). As most outer membrane proteins probably have a β-sheet structure (Vogel and Jähnig, 1986), this suggests a structure for AlkL in which the polypeptide traverses the membrane about 12 times, indicated by the numbers above regions separating the putative turns.

Growth of Pseudomonas putida *strains on ethylbenzene*

Fukuda *et al.* (1989) have reported that *P. oleovorans* GPo1 is able to grow on ethylbenzene, and that at least the catalytic component of the alkane hydroxylase system; AlkB, is required for growth on ethylbenzene. It is possible that the *alk* system has evolved to allow its host to grow on compounds other than alkanes, e.g. ethylbenzene. We therefore tested whether AlkJ, AlkK and AlkL are required for growth on ethylbenzene. Apparently this is not the case; plasmid pGEc41 allows *P. putida* GPo12 to grow as well on ethylbenzene as pGEc47.

Expression of AlkJ, AlkK and AlkL in Pseudomonas putida *and* Escherichia coli

To test whether AlkJ, AlkK and AlkL are expressed at significant levels, we fractionated *P. putida* and *E. coli* cells carrying plasmids pGEc47 and its deletion derivatives (Fig. 8, panel A). Lanes 1 and 2 show total membranes of *P. putida* GPo12 carrying pGEc47, uninduced and induced, respectively. Two major and two minor dicyclopropylketone-inducible bands are indicated by arrows. The 40 kDa band represents AlkB. The second major band has a molecular mass of about 59 kDa. This size suggested that it represented AlkJ or AlkK. In lanes 3, 4 and 5 total membranes of GPo12 carrying pGEc47ΔJ, pGEc47ΔK and pGEc41, respectively, are shown. In lanes 3 and 5 the 59 kDa band is not present, which indicates that the band represents AlkJ. The *N*-terminal sequence of the AlkJ band was determined, which resulted in the following sequence: Met–Tyr–Lys or Asp–Tyr–Val or Ile–Ile or Lys–Val–Gly–Ala–Gly. The underlined residues represent the first 10 amino acids of the translation of the *alkJ* ORF, confirming the identification of the 60 kDa band.

The *E. coli* equivalent of AlkK; FadD, is a membrane-bound protein, which suggests that AlkK is membrane bound as well. However, it is the cytoplasmic fraction of GPo12 (pGEc47) that contains a 60 kDa protein (Fig. 8, panel B, lane 2), absent when the *alk* genes are not induced (lane 1), and absent when part of the *alkK* gene is deleted (lane 3).

The primary sequence of AlkL suggests that it is localized in the outer membrane. We studied the expression of AlkL in *E. coli* strains, because *E. coli* membranes can easily be separated by sucrose-density centrifugation, which is not the case for *P. putida* membranes. *E. coli* W3110 was used because it expresses the Alk proteins to high levels (M. Nieboer and B. Witholt, unpublished). Panel C of Fig. 8 shows that in outer membranes of W3110 (pGEc47) cells, induced with dicyclopropylketone,

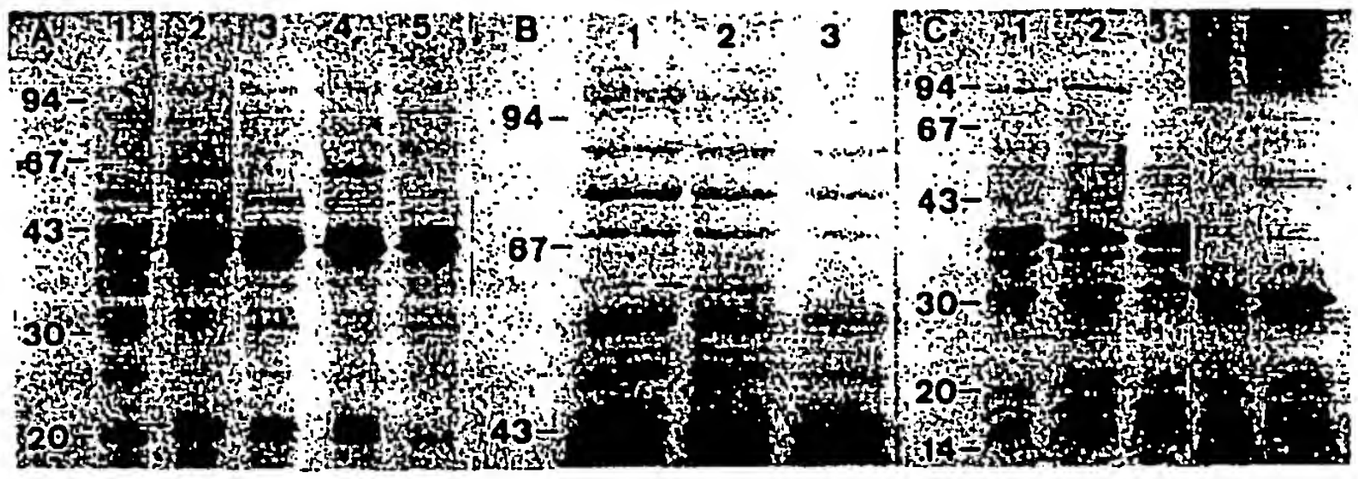

Fig. 8. Expression of AlkJ, AlkK and AlkL in *P. putida* and *E. coli*. *P. putida* GPo12 and *E. coli* W3110 strains were grown in E-2 medium and induced with DCPK for expression of the *alk* genes where specified, as described in the *Experimental procedures*. Cell fractions were isolated and analysed for the presence of AlkJ, AlkK and AlkL. Arrows mark dicyclopropylketone-inducible proteins.
A. Total membranes of *P. putida* GPo12 strains, incubated for 30 min at 37°C with 1 volume of denaturation buffer, and separated on a 12.5% polyacrylamide gel. Lane 1, GPo12 (pGEc47; lane 2, GPo12 (pGEc47) induced; lane 3, GPo12 (pGEc47ΔJ) induced; lane 4, GPo12 (pGEc47ΔK) induced: lane 5, GPo12 (pGEc41) induced.
B. Cytoplasmic fractions of *P. putida* GPo12 strains, incubated for 5 min at 100°C, and separated on a 7.5% polyacrylamide gel. Lane 1, GPo12 (pGEc47); lane 2, GPo12 (pGEc47) induced; lane 3, GPo12 (pGEc47ΔK) induced.
C. Outer membranes of *E. coli* W3110 strains, separated on a 12.5% gel. Samples shown in lanes 1, 2 and 3 were boiled, samples shown in lanes 4 and 5 were incubated at 37°C. Lane 1, W3110 (pGEc47); lane 2, W3110 (pGEc47) induced; lane 3, W3110 (pGEc47ΔL) induced; lane 4, W3110 (pGEc47); lane 5, W3110 (pGEc47) induced.

a 25 kDa band is present (lane 2), which is absent in outer membranes of W3110 (pGEc47ΔL) (lane 1), and uninduced W3110 (pGEc47) (lane 3). This strongly suggests that the band represents AlkL.

A 21 kDa dicyclopropylketone-inducible band is also visible in lanes 2 (pGEc47), 3 (pGEc47ΔJ) and 4 (pGEc47ΔK) of Fig. 8., Panel A. In all three cases the *alkL* gene is present. In lane 5 (pGEc41:ΔJKL) the band is absent, as would be expected if it represents AlkL. The size difference can, in part, be explained by the sample preparation. Total membranes of GPo12 are normally mixed 1:1 with denaturing buffer, and incubated at 37°C for 30 min, while *E. coli* outer-membrane preparations are usually boiled. When W3110 (pGEc47) outer membrane samples were treated at 37°C, the AlkL band shifted to 20 kDa (Panel C, lanes 4 and 5, uninduced and induced, respectively). Many other outer membrane proteins show the same heat-modifiable behaviour, as has been noted previously (Wensink and Witholt, 1981).

Discussion

In this paper we present the nucleotide sequence of the distal three genes of the *alkBFGHJKL* operon. We show that the *alkJ* gene encodes the NAD(P)-independent alcohol dehydrogenase studied by Benson and Shapiro (1976), *alkK* encodes an acyl-CoA synthetase and *alkL* encodes an outer membrane protein of unknown function. The results confirm the alcohol dehydrogenase marker-rescue experiments of Owen *et al.* (1984), and the peptide mapping by Eggink *et al.* (1987a). The work presented here, and that published before is summarized in Fig. 9.

Product of the alkJ *cistron*

The translation product of the *alkJ* cistron has previously been identified in *E. coli* minicell expression experiments (Eggink *et al.*, 1987a). The size and location of the first

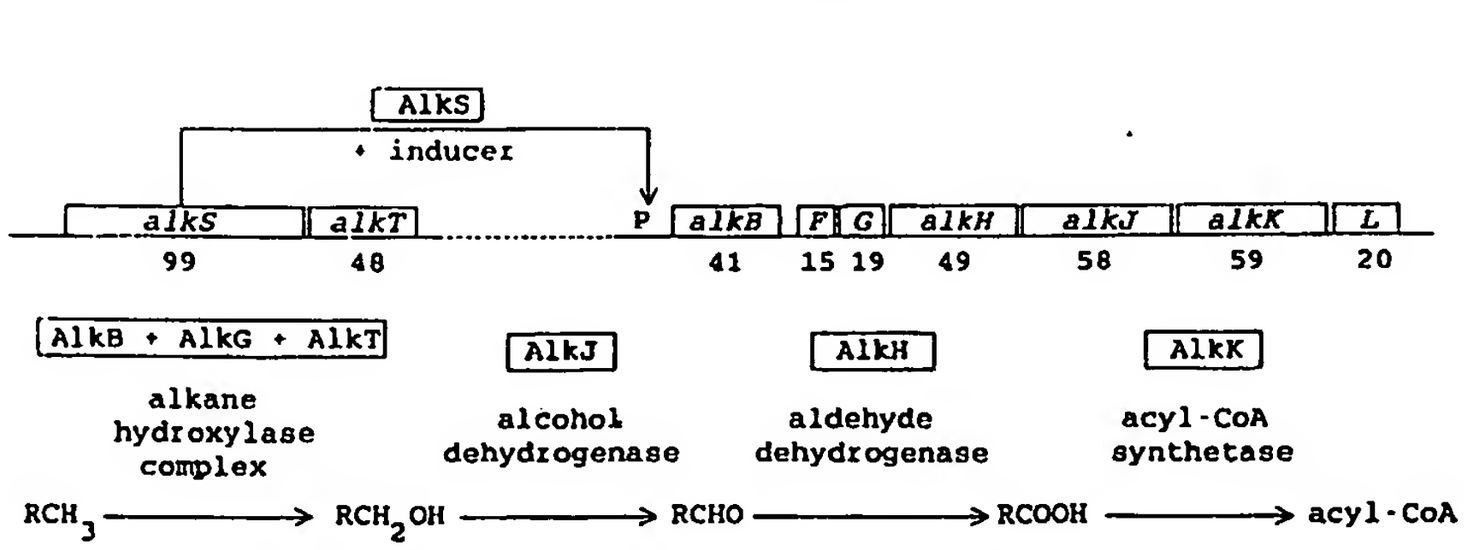

Fig. 9. Genetic structure of the *P. oleovorans alk* regulon and functions of the encoded proteins. The location of the *alk* open reading frames (open bars), the apparent molecular masses ($\times 10^{-3}$) of the Alk proteins determined in minicell expression experiments (Eggink *et al.*, 1987a), and their functions are shown.

ORF in the DNA sequence presented in Fig. 2 corresponds perfectly to that determined in these earlier experiments (Table I). The *N*-terminal sequence, as determined by gas-phase sequencing, was identical to that predicted from the DNA sequence.

The AlkJ amino acid sequence shares clear homology with a number of enzymes: choline dehydrogenase, glucose dehydrogenase, glucose oxidase and alcohol oxidase (Fig. 3), the last three of which have been shown to be flavin enzymes. AlkJ is NAD(P)-independent, and possesses a characteristic ADP-binding fingerprint close to the amino terminus, like the homologous enzymes (Fig. 5). This suggests that AlkJ binds FAD as well.

We could not detect any H_2O_2 production by strains carrying the *alkJ* gene. The possibility that AlkJ, like the choline dehydrogenase (BetA) is an electron transfer-dependent dehydrogenase was tested. The results show that oxygen is consumed when 1-octanol is added to membrane preparations containing AlkJ. The same preparations contain the PMS/DCPIP oxidizing activity studied by Benson and Shapiro (1976). Both activities are lost when a part of the *alkJ* gene is deleted.

The alignment of AlkJ in Fig. 3. shows that the dehydrogenases are related to the oxidases. In this respect it is interesting that the two oxidases have no access to the electron-transport chain. Yeast alcohol oxidase is located in the yeast peroxisome, while glucose oxidase is exported to the medium. *Drosophila* glucose dehydrogenase is exported into the ejaculatory ducts and has no access to the electron-transfer chain either. The *in vivo* electron acceptor of this enzyme is unknown (Cavener and MacIntyre, 1983).

The properties of AlkJ and BetA are similar to those of the dye-linked alcohol dehydrogenase of *P. aeruginosa* described by Tassin *et al.* (1973). The *P. aeruginosa* enzyme was later shown to be a periplasmic pyrroloquinoline quinone-containing enzyme (Groen *et al.*, 1984), linked to the electron-transfer chain through a special-purpose newly characterized cytochrome (J. A. Duine, personal communication). AlkJ and BetA may use coenzyme Q as their entry point for the electron-transfer chain, similar to the succinate, fumarate, lactate and NADH dehydrogenases, which are part of central metabolic pathways, contain FAD and can be assayed with PMS/DCPIP.

Product of the alkK *cistron*

No functions had previously been assigned to, or proposed for, the distal two cistrons of the operon because no mutations affecting alkane catabolism had been mapped to these genes. In addition, the alkane hydroxylase system, encoded by *alkB*, *alkG* and *alkT*, the alcohol dehydrogenase AlkJ and the aldehyde dehydrogenase AlkH together already constitute a complete pathway from alkanes to fatty acids.

The homology of AlkK to several acyl-CoA synthetases suggested that AlkK might catalyse the first step in the fatty acid β-oxidation pathway to form an alkanoyl-CoA intermediate. This notion was confirmed by complementation of an *E. coli* acyl-CoA synthetase (*fadD*) mutation by plasmids containing the *alkK* gene. *E. coli* K27 *fadR* carrying the *alk* system grows equally well on octane when compared with *E. coli* GEc137 carrying the *alk* system, which shows that *alkK* can fully complement the function of *fadD*. As AlkK only complements the *fadD* mutation for medium chain-length substrates we conclude that AlkK is specific for these substrates. The observation that AlkK is a soluble protein, unlike its *E. coli* equivalent FadD, also suggests that its substrates must be relatively soluble.

Product of the alkL *cistron*

The last ORF of the *alkBFGHJKL* operon (*alkL*) encodes a 25.0 kDa polypeptide, with a signal peptide-like sequence. This may be cleaved off to form the mature *alkL* gene product, homologous to a *V. cholerae* outer-membrane protein, OmpW. Several features of the primary structure suggest that AlkL is an outer membrane protein which could cross the membrane about 12 times in an amphipathic β-sheet structure. Outer membranes of W3110 carrying pGEc47 contain an alkane-inducible protein with a molecular mass of 25 kDa. When the *alkL* gene is not induced, or deleted, this protein band is not present.

Function of AlkJ, AlkK in alkane catabolism

AlkJ and AlkK were identified as alcohol dehydrogenase and acyl-CoA synthetase, respectively. AlkJ is essential for growth on alkanes only in *P. putida* alcohol dehydrogenase mutants, while AlkK is only essential for growth on alkanes when the corresponding chromosomal function is missing, e.g. in the *E. coli fadD* mutant K27. We do not know whether only the first step of fatty acid degradation is encoded by the *alk* operons. It is quite possible that other non-essential β-oxidation genes are located outside the regions that were cloned from the OCT-plasmid.

As AlkL is located in the outer membrane, it could be involved in transport of substrates. Many compounds that are oxidized by the alkane hydroxylase system *in vitro* (McKenna and Coon, 1970), are not converted by the enzyme system *in vivo* (Bosetti *et al.*, 1992; J. B. Van Beilen, unpublished results). Compounds that *are* converted *in vivo* may utilize a transport system to enter the cell. Experiments to test whether AlkL has a transport function have thus far yielded contradictory results (not

shown). However, in these experiments the alkane substrates.were present at 2% of the total culture volume. AlkL may be necessary only at (very) low substrate concentrations. An indication that this is the case is the low *Km* (less than 5 μM) of the alkane hydroxylase for *n*-octane. Alternatively, it is possible that the function of AlkL is specific for the original host (probably a Gram-negative bacterium because AlkL is an outer membrane protein), or for the original substrate of the *alk* proteins, which may not be a medium-chain-length alkane. The observation that the *alkJKL* genes are not necessary for growth on ethylbenzene suggests that this 'native' substrate may again be a different compound.

Evolutionary aspects of alkane oxidation

We have shown that the *P. oleovorans alk* system encodes several functions (AlkH, AlkJ and AlkK) that have a chromosomal equivalents (Kok *et al.*, 1989b), and two proteins the functions of which are unknown (AlkF and AlkL). The codon usage and G+C content of the *alk* system differ significantly from the rest of the OCT-plasmid and the *P. oleovorans* chromosome, suggesting a transfer of the *alk* system to *P. oleovorans* and the OCT-plasmid. The fact that the ORFs of the non-essential Alk polypeptides mentioned above are still intact indicates that this must have been a relatively recent event.

Experimental procedures

Bacterial strains, plasmids and growth conditions

The bacterial strains and plasmids used are listed in Table 2. Cells were grown either on tryptone–yeast extract (TYE) medium or on E-2 medium (Eggink *et al.*, 1987b), supplemented with carbon sources, thiamine or antibiotics. For growth on *n*-octane or 1-octanol, the Petri dishes were incubated at 32°C in a sealed container under *n*-octane or 1-octanol vapour. Ampicillin was used at 50 mg l^{-1}, tetracycline was used at 12.5 mg l^{-1}.

Mobilization of pLAFR1 derivatives from *E. coli* to *Pseudomonas* was done according to Figurski and Helinski (1979), by the replica plating technique using the helper plasmid pRK2013 (Ditta *et al.*, 1980). *E. coli* strains were cured of pGEc47 or its derivatives by growing cells in the absence of tetracycline while inducing the *alk* system with dicyclopropylketone.

DNA manipulations

Plasmid DNA was extracted according to Birnboim and Doly (1979). DNA fragments were isolated as described by Zhu *et al.* (1985). Restriction enzymes and T4 DNA ligase were purchased from Boehringer Mannheim and used according to the specifications of the supplier. *E. coli* strains were made competent for DNA transformations according to Chung and Miller (1988). Nucleotide sequencing was done according to Sanger *et al.* (1977) using M13mp18 and M13mp19 as vectors to prepare single-stranded DNA. The sequence strategy (Fig. 1E) was a combination of primer walking, shotgun cloning, and forced cloning.

The nucleotide and amino acid sequences were analysed using programs collected in the PC/GENE software package, developed by A. Bairoch (Genofit/Intelligenetics), the *NAQ* and *PSQ* database search programs from the Protein Identification

Table 2. List of strains and plasmids, and relevant genotype and/or phenotype.

Strains/ Plasmid	Relevant genotype or plasmid	Growth on octane	Source or Reference
E. coli			
SF800	W3110, *polA*		Laboratory collection
W3110	λ⁻, *thyA36*, *deoC2*, IN*1*		Bachmann (1987)
GEc137	DH1, *fadR*	–	Eggink *et al.* (1987b)
GEc93	GEc137, pGEc47	++++	Eggink *et al.* (1987b)
GEc139	GEc137, pGEc41	++	Eggink *et al.* (1987b)
GEc350	GEc137, pGEc47ΔJ	++	This study
GEc358	GEc137, pGEc47ΔK	+++	This study
GEc359	GEc137, pGEc47ΔL	+++	This study
K27	K12Ymel, *fadD*	–	Overath *et al.* (1969)
GEc351	K27, pGEc47	–	This study
GEc354	K27, *fadR*	–	This study
GEc355	GEc354, pGEc47	++++	This study
GEc356	GEc354, pGEc41	–	This study
GEc357	GEc354, pGEc47ΔJ	++	This study
GEc360	GEc354, pGEc47ΔK	–	This study
GEc361	W3110, pGEc47		This study
GEc362	W3110, pGEc47ΔL		This study
P. putida			
GPo1	Prototroph (OCT-1)	++++	Schwartz and McCoy (1973)
GPo12	GPo1 cured of OCT	–	Kok (1988)
GPp202	GPo12, pGEc47	++++	Eggink *et al.* (1987b)
GPp203	GPo12, pGEc41	++	Eggink *et al.* (1987b)
GPp204	GPo12, pGEc47ΔJ	++	This study
GPp205	GPo12, pGEc47ΔK	+++	This study
GPp206	GPo12, pGEc47ΔL	+++	This study
PpS81	PpG1 *alcA81*	–	Grund *et al.* (1975)
GPp10	PpS81 pGEc47	+++	Eggink *et al.* (1987b)
GPp11	PpS81 pGEc41	+/–	Eggink *et al.* (1987b)
GPp207	PpS81 pGEc47ΔJ	+/–	This study
GPp597	PpG1 *alcA437 his597*	–	Benson *et al.* (1987b)
GPp208	PpS597 pGEc47	+++	This study
GPp209	PpS597 pGEc41	+/–	This study
GPp210	PpS597 pGEc47ΔJ	+/–	This study
Plasmids			
pRK2013	Km, Tra, ColE1 replicon		Ditta *et al.* (1980)
pGEc51	pBR322, *alkHJK' Hin*dIII insert		Eggink *et al.* (1987a)
pGEc47	pLAFR1, *alkST/alkBFGHJKL*		Eggink *et al.* (1987b)
pGEc41	pLAFR1, *alkST/alkBFGH*		Eggink *et al.* (1987b)
pGEc51ΔJ	pGEc51, deletion in *alkJ*		This study
pUC18JLΔK	pUC18, see text and pGEc47ΔK		This study
pUC18JKΔL	pUC18, see text and pGEc47ΔL		This study
pGEc47ΔJ	pGEc47, 774 bp *Ps*ti deletion in *alkJ*		This study
pGEc47ΔK	pGEc47, 568 bp *Hin*dIII deletion in *alkK*		This study
pGEc47ΔL	pGEc47, deletion of *alkL* from *Sal*(8000) to *Bam*HI (~9900)		This study

Resource (Dayhoff *et al.*, 1983) on the VAX11/750 computer, the Swiss-Prot Protein database (EMBL, Heidelberg, release 18), and the FASTP sequence alignment program (Lipman and Pearson, 1985). Multiple alignments and evolutionary trees were generated by the method of Higgins and Sharp (1988).

N-terminal sequence determination

Proteins separated by SDS–polyacrylamide gelelectrophoresis were transferred to polyvinylidene difluoride (PVDF) membranes (Immobilon Transfer, Millipore) according to Matsudaira (1987). The protein bands were stained with Coomassie brilliant blue and cut from the membrane. The *N*-terminal amino acid sequence was determined by Eurosequence B.V., Groningen, The Netherlands with a Model 477A, Applied Biosystems gas-phase protein sequenator.

Construction of deletion derivatives of pGEc47

Plasmids pGEc47ΔJ, pGEc47ΔK, pGEc47ΔL are derivatives of pGEc47 in which a 747 bp *Pst*I fragment was deleted from the *alkJ* gene, a 586 bp *Hin*dIII fragment was deleted from the *alkK* gene, and a ≈1900 bp *Sal*I–*Bam*HI fragment containing the *alkL* gene was deleted, respectively. All three deletions were first created in pMB1 plasmids carrying only (part of) the *alkJKL* region. The deletions were then transferred to pGEc47 by marker rescue as described below.

To construct pGEc47ΔJ a *Pst*I fragment (4583–5357) was deleted from pGEc51 (Eggink *et al.*, 1987a), which contains a *Hin*dIII fragment (2753–7337) with *alkH*, *alkJ*, and part of *alkK*. For *alkK* the *Pst*I(5357)–*Hin*dIII(7337) and *Hin*dIII(7923)-*Eco*RI(≈10500) fragments were combined with pUC18 cut with *Pst*I and *Eco*RI. This resulted in a plasmid which contains the entire *alkJKL* region, except for part of *alkK*. To construct an *alkL* deletion, first the *Pst*I(5357)–*Sal*I(8000) fragment was cloned in pUC18, and subsequently the *Bam*HI(≈9900)–*Eco*RI(≈10500) fragment was added, resulting in a construct which lacks *alkL* but contains the flanking DNA (a simple *Sal*I deletion was not possible because of the presence of a previously unreported *Sal*I site close to the *Eco*RI site).

E. coli SF800 (polA–) containing pGEc47 was transformed with the three deletion plasmids, pGEc51ΔJ, pUC18JLΔK, and pUC18JKΔL. Because these plasmids are pMB1 plasmids they cannot replicate in SF800. Ampicillin-resistant (Ap^R) colonies can only be obtained by homologous recombination of these plasmids into pGEc47. A subsequent selection for loss of the Ap marker, again by homologous recombination, effectively resulted in transfer of the deletions to pGEc47 in about 50% of the ampicillin-sensitive (Ap^S) colonies.

Alcohol dehydrogenase assay

E. coli and *P. putida* recombinants were grown overnight in 50 ml LB medium. These cultures were used to inoculate 500 ml LB medium. After 1 h of incubation at 30°C, the *alk* genes were induced by the addition of dicyclopropylketone to 0.02%. The cultures were harvested 3 h after induction. Total membrane preparations for alcohol dehydrogenase assays were prepared as described below. Protein concentrations were determined according to Markwell *et al.* (1981). The alcohol dehydrogenase assay based on the reduction of DCPIP in the presence of PMS as described previously (Tassin *et al.*, 1973). The assay system contained 2.7 ml of 50 mM Tris-HCl (pH 7.5), 1 mM KCN, 2.2 mM 1-octanol and 0.2 mM DCPIP. This solution was allowed to equilibrate for 5 min at 30°C before 0.1 ml of a 2 mg ml^{-1} PMS solution was added. After a blank rate was recorded appropriate quantities of membranes (10–50 µl, 0.1–1.0 mg of protein) were added to initiate the reaction. The activity was assayed by measuring the decrease in absorbance of DCPIP at 600 nm. Activity is expressed as micromoles of DCPIP reduced per minute per milligram of protein. DCPIP has an extinction coefficient of 20.6 mM^{-1} cm^{-1} at 600 nm.

Oxygen consumption by membrane preparations was measured using a Biological Oxygen Monitor. To a 5 ml volume of 50 mM Tris-HCl buffer pH 7.4, saturated with oxygen, 0.5 ml of membrane suspension was added. This mixture was allowed to equilibrate for 2 min. 1-Octanol was added to 2 mM as a 1 M solution in ethanol. KCN was added as a 1 M solution in 50 mM Tris-HCl buffer, pH 8.0. PMS was added as a 2 mg ml^{-1} solution in 50 mM Tris-HCl buffer pH 7.4. To test NADH oxidase activity NADH was added to 100 µg ml^{-1}. Triton X-100 was used at a concentration of 1 %.

Alcohol oxidase assay

Alcohol oxidase activity was assayed according to Verduyn *et al.* (1984). The chromogenic substrate ABTS is converted into a green product by peroxidase and hydrogen peroxide. The assays were carried out in 50 mM K-phosphate pH 7.5, 1 mM Naazide. Azide was added to prevent interference from catalase activity. The AlkJ and peroxidase activities were not affected by the addition of azide, as tested for AlkJ in the Biological Oxygen Monitor, and for peroxidase by measuring the disappearance of H_2O_2 at 240 nm. Calculations were based on an extinction coefficient of the green product of 43.2 $mM^{-1}cm^{-1}$ at 412 nm.

Isolation of membrane fractions

P. putida total membranes were isolated essentially as described by Van Heerikhuizen *et al.* (1975). 2 ml L-broth precultures were inoculated from fresh plates and grown during daytime. Erlenmeyer flasks containing 50 ml of E2-medium with 0.5 % glucose were inoculated with 2 ml of the preculture. After 2 h the cultures were induced with 10 µl dicyclopropylketone. Cells were harvested after overnight incubation at 30°C, and converted to sphaeroplasts. These were disrupted by passage through a French Press cell at 69000 kPa. After removal of large fragments and whole cells, total membranes were pelleted. *E. coli* membranes were separated on a sucrose-density gradient. Samples were analysed on 7.5 or 12.5 % SDS–polyacrylamide gels (Laemmli, 1970).

Acknowledgements

We thank Peter Terpstra for helpful discussions, and providing support in analysing DNA and protein sequences. This

research was supported by the Programma Commissie voor Biotechnologie (PCB) of the Netherlands Ministry of Economic Affairs.

References

Bachmann, B.J. (1987) Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12. In: Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology, Vol. 2.* Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., and Umbarger, H.E. (eds.) Washington D.C.: American Society for Microbiology pp. 1190–1219.

Bairoch, A. (1992) PROSITE: a dictionary of sites and patterns in proteins. *Nucl Acids Res* **20:** 2013–2018. Release 8.10 of PROSITE, March 1992.

Benson, S., and Shapiro, J. (1976) Plasmid determined alcohol dehydrogenase activity in alkane-utilizing strain of *Pseudomonas putida. J Bacteriol* **126:** 794–798.

Benson, S., Fennewald, M., Shapiro, J., and Huettner, C. (1977) Fractionation of inducible alkane hydroxylase activity in *Pseudomonas putida* and characterization of hydroxylase-negative plasmid mutations. *J Bacteriol* **132:** 614–621.

Birnboim, H.C., and Doly, J. (1979) A rapid alkaline extraction procedure for screening of recombinant plasmid DNA. *NuclAcids Res* **7:** 1513–1523.

Bosetti, A., Van Beilen, J.B., Preusting, H., Lageveen, R.G., and Witholt, B. (1992) Production of primary aliphatic alcohols with a recombinant *Pseudomonas* strain, encoding the alkane hydroxylase enzyme system. *Enzyme Microb Technol* **14:** 702–708.

Cavener, D.R., and MacIntyre, R.J. (1983) Biphasic expression and function of glucose dehydrogenase in *Drosophila melanogaster. Proc Natl Acad Sci USA* **80:** 6286–6288.

Chakrabarty, A.M., Chou, G., and Gunsalus, I.C. (1973) Genetic regulation of octane dissimulation plasmid in *Pseudomonas. Proc Natl Acad Sci USA* **70:** 1137–1140.

Chung, C.T., and Miller, R. H. (1988) A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nucl Acids Res* **16:** 3580.

Dayhoff, M.O., Barker, W.C., and Hunt, L.T. (1983) Establishing homologies in protein sequences. *Methods Enzymol* **91:** 524–545.

Ditta, G., Stanfield, S., Corbin, D., and Helinski, D.R. (1980) Broad host range DNA cloning system for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti. Proc Natl Acad Sci USA* **77:** 7347–7351.

Eggink, G., van Lelyveld, P.H., Arnberg, A., Arfman, N., Witteveen, C., and Witholt, B. (1987a) Structure of the *Pseudomonas putida alkBAC* operon. Identification of transcription and translation products. *J Biol Chem* **262:** 6400–6406.

Eggink, G., Lageveen, R.G., Altenburg, B., and Witholt, B. (1987b) Controlled and functional expression of the *Pseudomonas oleovorans* alkane utilization system in *Pseudomonas putida* and *Escherichia coli. J Biol Chem* **262:** 17712–17718.

Eggink, G., Engel, H., Meijer, W., Otten, J., Kingma, J., and Witholt, B. (1988) Alkane utilization in *Pseudomonas oleovorans.* Structure and function of the regulatory locus *alkR. J Biol Chem* **263:** 13400–13405.

Eggink, G., Engel, H., Vriend, G., Terpstra, P., and Witholt, B. (1990) Rubredoxin reductase of *Pseudomonas oleovorans.* Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. *J Mol Biol* **212:** 135–142.

Higgins, D.G., and Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73:** 237–244.

Fennewald, M., Prevatt, W., Meyer, R., and Shapiro. J. (1978) Isolation of Inc P-2 plasmid DNA from *Pseudomonas aeruginosa. Plasmid* **1:** 164–173.

Fennewald, M., Benson, S., Oppici, M., and Shapiro, J. (1979) Insertion element analysis and mapping of the *Pseudomonas* plasmid *alk* regulon. *J Bacteriol* **139:** 940–952.

Figurski, D.H., and Helinski, D.R. (1979) Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci USA* **76:** 1648–1652.

Frederick, K.R., Tung, J., Emerick, R.S., Masiarz, F.R., Chamberlain, S.H., Vasavada, A., Rosenberg, S., Chakraborty, S., Schopter, L.M., and Massay, V. (1990) Glucose oxidase from *Aspergillus niger.* Cloning, gene sequence, secretion from *Saccharomyces cerevisiae* and kinetic analysis of a yeast-derived enzyme. *J Biol Chem* **265:** 3793–3802.

Fukuda, M., Nishi, T., Igarashi, M., Kondo, T., Takagi, M., and Yano, K. (1989) Degradation of ethylbenzene by *Pseudomonas putida* harboring OCT plasmid. *Agric Biol Chem* **53:** 3293–3299.

Groen, B.W., Frank, J., and Duine, J.A. (1984) Quinoprotein alcohol dehydrogenase from ethanol-grown *Pseudomonas aeruginosa. Biochem J* **223:** 921–924.

Grund, A., Shapiro, J., Fennewald, M., Bacha, P.. Leahy, J., Markbreiter, K., Nieder, M., and Toepfer, M. (1975) Regulation of alkane oxidation in *Pseudomonas putida. J Bacteriol* **123:** 546–556.

Jalajakumari, M.B., and Manning, P.A. (1990) Nucleotide sequence of the gene, *ompW,* encoding a 22 kDa immunogenic outer membrane protein of *Vibrio cholerae. Nucl Acids Res* **18:** 2180.

Kok, M. (1988) Alkane utilization by *Pseudomonas oleovorans.* Ph.D. thesis. University of Groningen. The Netherlands.

Kok, M., Oldenhuis. R.. van der Linden, M.P.G., Raatjes, P., Kingma, J., van Lelyveld, P.H., and Witholt, B. (1989a) The *Pseudomonas oleovorans* alkane hydroxylase gene. Sequence and expression. *J Biol Chem* **264:** 5435–5441.

Kok, M., Oldenhuis, R.. van der Linden, M.P.G., Meulenberg, C. H.C., Kingma, J., and Witholt, B. (1989b) The *Pseudomonas oleovorans alkBAC* operon encodes two structurally related rubredoxins and an aldehyde dehydrogenase. *J Biol Chem* **264:** 5442–5451.

Krasney, P.A., Carr, C.M., and Cavener, D.R. (1990) Evolution of the glucose dehydrogenase gene in *Drosophila. Mol Biol Evol* **7:** 155–177.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680–685.

Lamark, T., Kaasen, I., Eshoo, M.W., Falkenberg, P., McDougal, J. and Strom, A.R. (1991) DNA sequence and analysis of the *bet* genes encoding the osmoregulatory choline–glycine betaine pathway of *Escherichia coli. Mol Microbiol* **5:** 1049–1064.

Ledeboer, A.M., Edens, L., Maat, J., Visser, C., Bos. J.W., and Verrips, C.T. (1985) Molecular cloning and characterization of a gene coding for methanol oxidase in *Hansenula polymorfa. Nucl Acids Res* **13:** 3063–3082.

Lipman, D.J., and Pearson, W.R. (1985) Rapid and sensitive protein similarity searches. *Science* **227:** 1435–1441.

McKenna, E.J., and Coon, M.J. (1970) Enzymatic ω-oxidation. IV. Purification and properties of the ω-hydroxylase of *Pseudomonas oleovorans*. *J Biol Chem* **245**: 3882–3889.

Mandel, M. (1966) Deoxyribonucleic acid base composition in the genus *pseudomonas*. *J Gen Microbiol* **43**: 273–292.

Markwell, M.A.K., Haas, S.M., Tolbert, N.E., and Bieber, L.L. (1981) Protein determination in membrane and lipoprotein samples: Manual and automated procedures. *Methods Enzymol* **72**: 296–298.

Matsudaira, P. (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J Biol Chem* **262**: 10035–10038.

Overath, P., Pauli, G., and Schairer, H. (1969) Fatty acid degradation in *Escherichia coli*. An inducible acyl-CoA synthetase, the mapping of *old*-mutations, and the isolation of regulatory mutants. *Eur J Biochem* **7**: 559–574.

Owen, D.J., Eggink, G., Hauer, B., Kok, M., McBeth, D.L., Yang, Y.L., and Shapiro, J.A. (1984) Physical structure, genetic content and expression of the *alkBAC* operon. *Mol Gen Genet* **197**: 373–383.

Paul, C., and Rosenbusch, J.P. (1985) Folding patterns of porin and bacteriorhodopsin. *EMBO J* **4**: 1593–1597.

Peterson, Basu and Coon, M.J. (1966) Enzymatic ω-oxidation. Electron carriers in fatty acid and hydrocarbon hydroxylation. *J Biol Chem* **241**: 5162–5164.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463–5467.

Saraste, M., Sibbald, P.R., and Wittinghofer, A. (1990) The P-loop — a common motif in ATP- and GTP-binding proteins. *TIBS* **15**: 430–434.

Scholten, J.D., Chang, K.-H., Babbitt, P.C., Charest, H., Sylvestre, M., and Dunaway-Mariano, D. (1991) Novel enzymic hydrolytic dehalogenation of a chlorinated aromatic. *Science* **253**: 182–185.

Schwartz, R.D., and McCoy, C.J. (1973) *Pseudomonas oleovorans* hydroxylation–epoxidation system: Additional strain improvements. *Appl Microbiol* **26**: 217–218.

Shine, J., and Dalgarno, L. (1975) Determinants of cistron specificity in bacterial ribosomes. *Nature* **254**: 34–38.

Stoltzfus, A., Leslie, J.F., and Milkman, R. (1988) Molecular evolution of the *Escherichia coli* chromosome. I. Analysis of structure and natural variation in a previously uncharacterized region between *trp* and *tonB*. *Genetics* **120**: 345–358.

Struyvé, M., Moons, M., and Tommassen, J. (1991) Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *J Mol Biol* **218**: 141–148.

Tassin, J.P., Calier, C., and Vandecasteele, J.P. (1973) Purification and properties of a membrane-bound alcohol dehydrogenase involved in oxidation of long-chain hydrocarbons by *Pseudomonas aeruginosa*. *Biochim Biophys Acta* **315**: 220–232.

Turgay, K., Krause, M., and Marahiel, M.A. (1992) Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol Microbiol* **6**: 529–546.

Van Heerikhuizen, H., Kwak, E., Van Bruggen, E.F.J., and Witholt, B. (1975) Characterization of a low density cytoplasmic membrane fraction isolated from *Escherichia coli*. *Biochim Biophys Acta* **413**:177–191.

Verduyn, C., Van Dijken, J. P., and Scheffers, W. A. (1984) Colorimetric alcohol assays with alcohol oxidase. *J Microbiol Methods* **2**: 15–25.

Vogel, H., and Jähnig, F. (1986) Models for the structure of outer membrane proteins of *Escherichia coli* derived from Raman spectroscopy and prediction methods. *J Mol Biol* **190**: 191–199.

Von Heyne, G. (1986) A new method for predicting signal sequence cleavage sites. *Nucl Acids Res* **14**: 4683–4690.

Wierenga, R.K., Terpstra, P., and Hol, W.G.J. (1986) Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. *J Mol Biol* **187**: 101–107.

Wensink, J., and Witholt, B. (1981) Identification of different forms of the murein-bound lipoprotein found in isolated outer membranes of *Escherichia coli*. *Eur J Biochem* **113**: 349–357.

Zhu, J., Kempenaers, W., Van der Straeten, D., Contreras, R., and Fiers, W. (1985) A method for fast and pure DNA elution from agarose gels by centrifugal filtration. *Bio/Technology* **3**: 1014–1016.

Article No. jmbi.1999.2964 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 291, 59–70

JMB

The Cloning and Expression of *Pf*acs1, a *Plasmodium falciparum* Fatty Acyl Coenzyme A Synthetase-1 Targeted to the Host Erythrocyte Cytoplasm

Fuencisla Matesanz, Isabel Durán-Chica and Antonio Alcina*

Instituto de Parasitología y Biomedicina "López Neyra" CSIC, Granada, Spain

**Corresponding author*

Plasmodium is unable to carry out *de novo* fatty acid synthesis and has to obtain these compounds from their host for subsequent activation by thioesterification with coenzyme A. This activity is catalyzed by a fatty acyl-CoA synthetase enzyme (EC 6.2.1.3). Here, we describe a novel gene from *P. falciparum* whose recombinant purified product from baculovirus-transfected insect cell line had the enzymatic activity of a long-chain fatty acyl-CoA synthetase. It was named *pf*acs1, since it belongs to a multi-member gene family as revealed by the sequence of several clones and a multi-band pattern in Southern blots. The sequence specifies a product of 820 amino acid residues. It was transcribed and expressed in infected erythrocytes having an apparent molecular mass of 100 kDa. Immuno-labeling of infected erythrocytes with a specific antibody against the carboxy-terminal part of the PfACS1 localized the product early after the erythrocyte invasion in vesicle-like structures budding off the parasitoforous membrane toward the red cell cytoplasm. Its unique carboxy-terminal structure of 70 extra amino acid residues, longer than any other reported acyl-CoA synthetase, is probably related to its localization in the cytoplasm of the host erythrocyte. The phylogenetic relationship among other AMP-forming enzymes, placed PfACS1 closer to *Saccharomyces cerevisiae*, sharing significant amino acid identities, especially in the conserved signature motif that modulates fatty acid substrate specificity and ATP/AMP-binding domains. Taking into account the importance of this enzymatic activity for the parasite, its extra-cellular location inside the infected erythrocyte, and the divergence with respect to the homologous human enzymes, it may be an important protein as a potential target candidate for chemotherapeutic antimalaria drugs.

Keywords: Plasmodium falciparum-infected erythrocyte; acyl-CoA synthetase; exported enzyme; immuno-labeling; baculovirus expression

Introduction

Malaria remains one of the most important diseases of humans in terms of both mortality and morbidity, with *Plasmodium falciparum* being the most important infecting agent, killing over one million children each year (Sturchler, 1989). Continual exposure of malarial parasite populations to different drugs may have selected for resistance to individual drugs and for genetic traits that favor initiation of resistance to novel unrelated antimalarials (Rathod *et al.*, 1997). Given this situation, there is a desperate need for new and better therapeutic drugs and targets.

During part of its life-cycle, the malaria parasite lives inside vertebrate erythrocytes. Although during its intraerythrocytic development *Plasmodium* spp causes a considerable increase in total fatty acid content of the host cell (Simoes *et al.*, 1992; Beaumelle & Vial, 1988), at this stage neither the parasite nor the host erythrocyte possesses biochemical pathways for the synthesis of fatty acids

Abbreviations used: ACS, acyl-CoA synthetase; CBB, Coomassie brilliant blue; CFA, complete Freund's adjuvant; CoA, Coenzyme A; FITC, fluorescein isothiocyanate; iE, *P. falciparum*-infected erythrocyte; IFA, immunofluorescence assay; PfACS1, *P. falciparum* acyl-CoA synthetase-1; RT-PCR, reverse transcription-polymerase chain reaction.

E-mail address of the corresponding author: pulgoso@ipb.csic.es

0022-2836/99/310059–12 $30.00/0

(Holz, 1977; Haldar *et al.*, 1985). To fulfil its nutritional requirements and construct a complex series of membranous tubules and vesicles throughout the erythrocyte, the parasite is able to actively take up fatty acids, either in free form from the blood plasma or as lysophospholipids from the host erythrocyte membrane, and to incorporate these into complex molecules, such as diacylphospholipids (Lauer *et al.*, 1997; Glick & Rothman, 1987). The first step for the incorporation of fatty acids is the thioesterification with coenzyme A (CoA), an enzyme activity that is known to be increased 20-fold in *Plasmodium knowlesi*-infected cells (Beaumelle & Vial, 1988). In mammals, fatty acid utilization is initiated after their activation catalyzed by acyl-CoA synthetases (ACS, fatty acid:CoA ligase, AMP-forming; EC 6.2.1.3). The acyl-CoA, produced from fatty acid, ATP, and CoA by ACS, is a key intermediate in various metabolic pathways including protein transport, enzyme activation, protein acylation, cell signaling, and transcriptional control, in addition to serving as substrates for β-oxidation and phospholipid biosynthesis (McLaughlin & Aderem, 1995; Korchak *et al.*, 1994).

In this work, we have isolated and characterized the first acyl-CoA synthetase from *P. falciparum* (PfACS1). It belongs to a multigene-member family with different isotypes. PfACS1 may be a candidate target enzyme for a chemotherapeutic approach, since it is essential for the parasite development, and there are large differences in the amino acid sequences between *P. falciaparum* enzymes and those reported for humans.

Results

Isolation and characteristics of the *Pf*acs1 gene

Our experiments were designed to complete the coding sequence of a *P. falciparum* antigen-encoding clone (clone 15) that showed significant homology to acyl-CoA synthetases from other organisms, and it was named PfACS1 in view of the existence of several homologous genes in the parasite genome (Figures 1 and 2). Toward this end, we isolated two overlapping λgt10 *P. falciparum* clones by screening a library first with the clone 15 probe (probe-c) obtaining the clone 34, and secondly, with a probe from the 5′ end of clone 34 (probe-a) resulting in the close 27. Two primers (IP19 and IP20) at the 3′ and 5′-ends of the presumed coding sequence were used to PCR amplify the complete insert from *P. falciparum* genomic DNA in three independent PCR reactions. Eight IP19-IP20 fragments were subcloned and sequenced. Some punctual nucleotide changes found in only one of these clones were considered PCR artifacts and the consensus sequence is represented in Figure 1. The longest open reading frame (ORF) of the *Pf*acs1 sequence was 2460 bp, potentially encoding a protein of 820 amino acid residues with a calculated molecular mass of

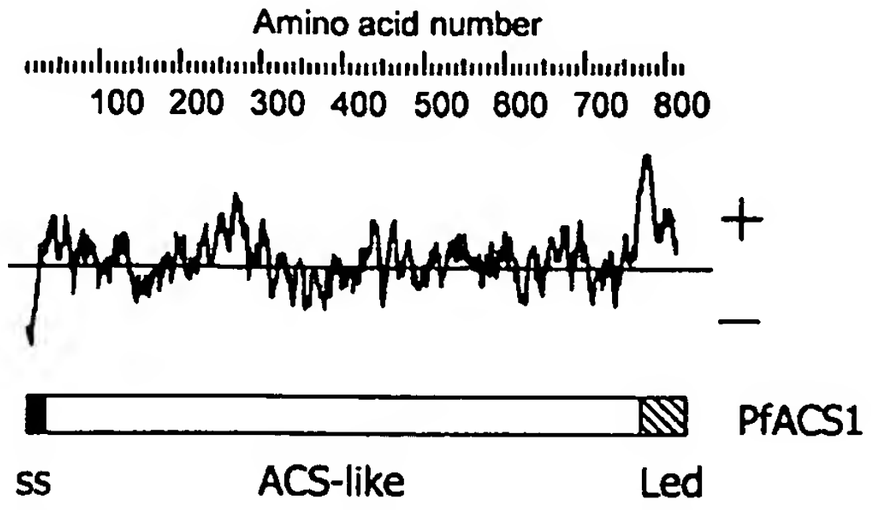

Figure 1. Schemes of general characteristics of the deduced amino acid sequence of the *Pf*acs1 gene. The hydropathy plot of the product according to Hopp & Woods (1981) is shown in which positive index represents hydrophilicity and negative index hydrophobicity. The scheme of the PfACS1 (*P. falciparum* acyl CoA synthetase-1), and the presumed relevant parts of the amino acid sequence are shown: ss (black box), stands for signal sequence; ACS-like (white box), stands for acyl-CoA synthetase-like sequence; and Led (striped box), signifies lysine-enriched domain. The GenBank accession number of *pf*acs1 is AF007828.

94,774 Da. The deduced amino acid sequence has an unusually high content of basic amino acids (15 Arg, 89 Lys, and 12 His) especially in the 70 carboxy-terminal residues, where the lysine content is 22 % (Lys-enriched domain, Led). Consequently, it has a deduced isoelectric point (pI) of 9.67. The N terminus (residues 1-20) has the characteristic features of a signal sequence (ss) with a hydrophobic core between residues 3 and 15, which is an essential part required for targeting and membrane insertion (Martoglio & Dobberrstein, 1998). The Led contains potential motifs for protein kinase C phosphorylation, *N*-myristoylation and *N*-glycosylation.

*Pf*acs1 homologous genes, transcription and expression of the product in *P. falciparum*

In order to determine the copy number of the *Pf*acs1 gene, Southern blot experiments were performed using the radiolabeled probes from the 5′-region (probe-a) and 3′-region (probe-b) of the *Pf*acs1 coding sequence that do not contain cleavage sites for any of the restriction enzymes used (Figure 2). Digestion of the DNA with *Eco*RI plus *Cla*I was clearly indicative of a multiple band pattern. Hybridization with probe-b revealed the existence of at least four homologous gene sequences. The probe-a reacted with only two of these four homologous sequences (Figure 2(a)). In spite of these apparent 5′-end polymorphisms, all of these homologous sequences contained two *Nde*I sites at approximately the same distance, giving rise to a 2.1 kb fragment (possibly containing two close bands in Figure 2(b), lanes 3 and 5). Therefore, these data indicate the existence of a multi-

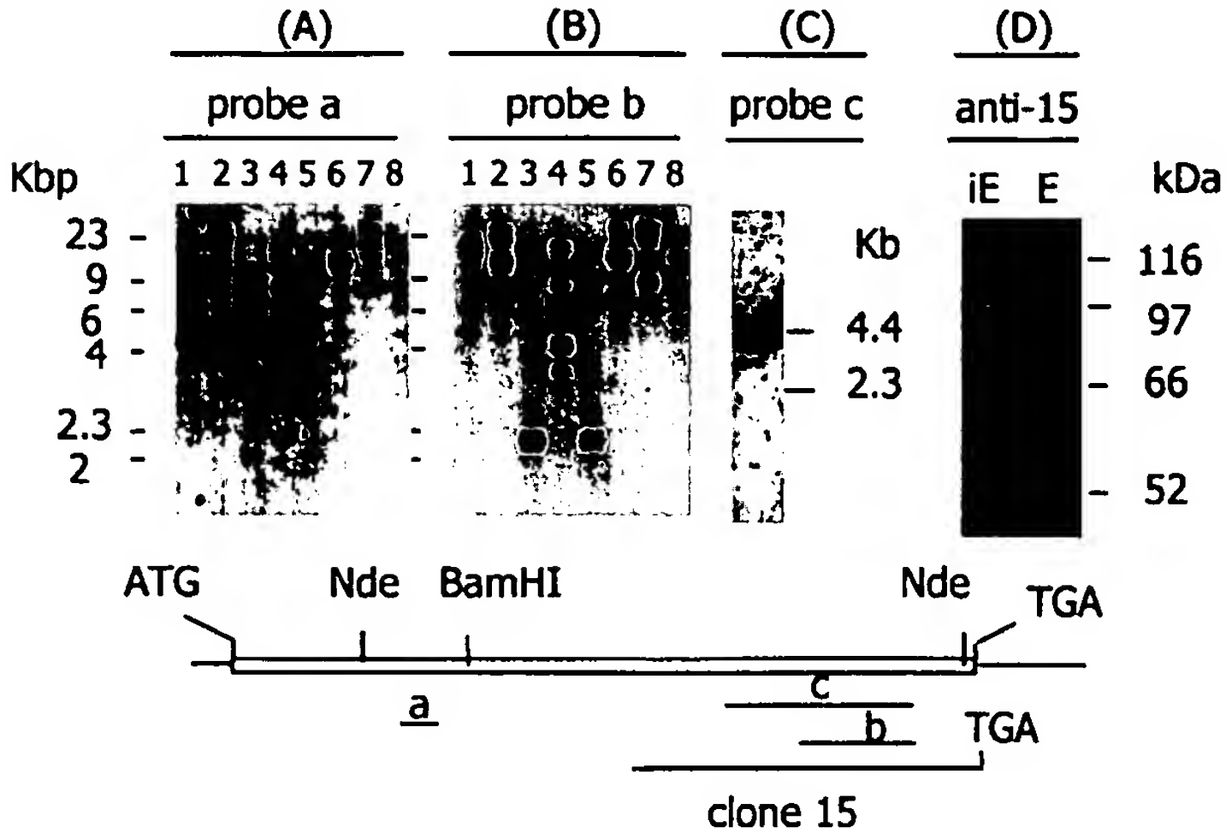

Figure 2. Southern, Northern and Western blot analysis of *pf*acs1. (a) and (b) show Southerns of *P. falciparum* genomic DNA probed with radiolabeled probe-a (IP15-IP18, 151 bp) and probe-b (1389-IP8, 434 bp) (see bottom scheme) as explained in Materials and Methods. DNA was digested with *Eco*RI (lane 1), *Cla*I (lane 2), *Nde*I (lane 3), *Eco*RI + *Cla*I (lane 4), *Eco*RI + *Nde*I (lane 5), *Hin*dIII (lane 6), *Bam*HI (lane 7), and *Bam*HI + *Hin*dIII (lane 8). (c) Northern blot of total RNA from *P. falciparum*-infected erythrocytes probed with DIG-labeled probe-c (1165-IP8, 715 bp) (see bottom scheme). (d) Western blots of extracts from saponin-treated *P. falciparum*-infected erythrocytes (iE) and saponin-treated uninfected erythrocytes (E), separated in SDS-10% PAGE, and incubated with the specific anti-15 antibody at 1/1000 dilution. Control Western blots with pre-immune serum, anti-TrpE antibody, and secondary antibody alone were negative. The exact position of the probes are indicated in Materials and Methods.

gene family comprising at least four members homologous but not identical with the *Pf*acs1 gene.

Analysis of total RNA from *P. falciparum*-infected erythrocytes by hybridization of blots with the 3′-region probe-c revealed a single transcript of about 3.5 kb (Figure 2(c)). The transcription of the *Pf*acs1 gene was confirmed by RT-PCR and sequencing (not shown) of total parasite RNA. Therefore, the *Pf*acs1 gene is transcribed in *P. falciparum*-infected erythrocytes.

To confirm the expression at the protein level and to determine the size of the PfACS1 gene product, Western blots of saponin-treated *P. falciparum*-infected erythrocyte extracts were performed (Figure 2(d)) using the anti-15 antibody (specific for the carboxy-terminal half of the PfACS1 protein). This antibody reacted with a major protein band with an apparent molecular mass of 100 kDa as determined by SDS-PAGE and a minor band with a molecular mass of around 75 kDa.

Comparison of PfACS1 with the GenBank data sequence

Searches of protein databases revealed significant similarities of PfACS1 with long-chain fatty acid CoA synthetases from bacterial to mammals (EC 6.2.1.3.). The homology search revealed that the PfACS1 sequence is more closely related to the acyl-CoA synthetase of yeast *Saccharomyces cerevisiae* than with other higher or lower eukaryotes and prokaryotes. PfACS1 seems to be closer to the yeast Faa1p, Faa3p and Faa4p acyl-CoA synthetases than Faa2p (Johnson *et al.*, 1994; Knoll *et al.*, 1994) (Figure 3). Sequence alignments indicated that the PfACS1 protein has 20-23% identity and 32-34% similarity with human, yeast rat, mouse, *Brassica napus* or *Haemophilus influenzae* fatty acid acyl-CoA synthetases (not shown). Further analysis of these data indicated that several regions of these enzymes were highly conserved (Figure 4). Several dispersed regions within the PfACS1 sequence seem very similar to others found in adenylate-forming enzymes containing an

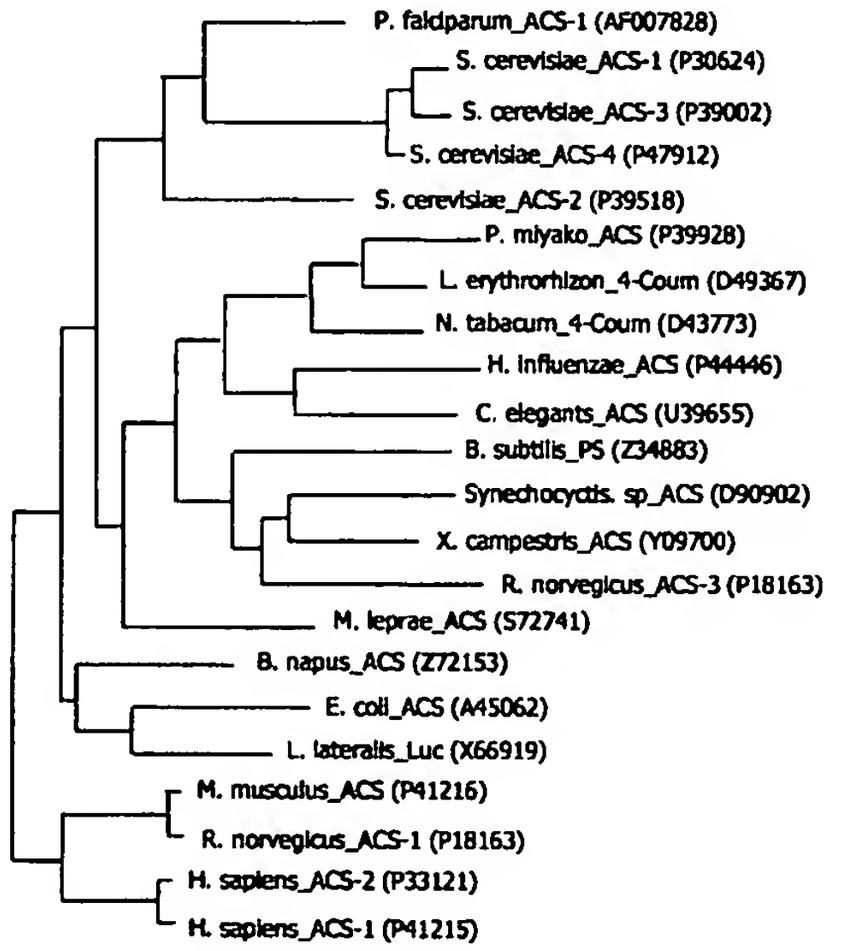

Figure 3. Phylogenetic analysis of PfACS1 and other members of the adenylate (AMP)-forming enzyme family. The tree was constructed by the GROWTREE and the DISTANCE programs of the GCG package and corrected using the Juker-Cantor method. Different acyl-CoA synthetases (ACS), peptide synthetases (PS) (Surfactine), firefly (*Photinus pyralis*) luciferase (Luc), and coumarate ligases (Coum) are shown.

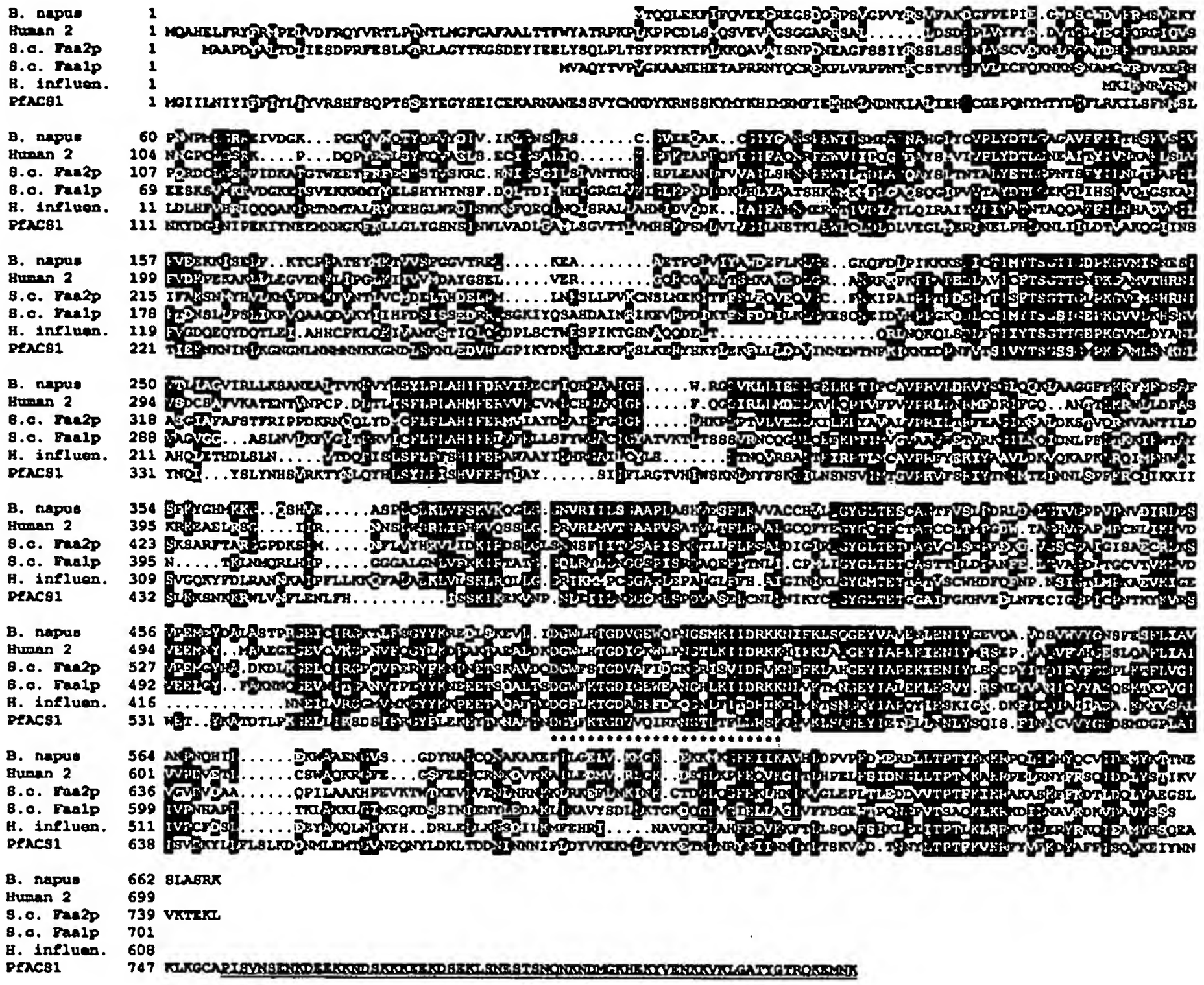

Figure 4. Alignment of the PfACS1 and other fatty acid CoA synthetases. *Brassica napus* (Z72153), Human 2 (P33121), *Saccharomyces cerevisiae* Faa1p (S. c. Faa1p) (P30624), S. c. Faa2p (P39518), and *Haemophilus influenzae* (P44446), *P. falciparum* acyl-CoA synthetase 1 (PfACS1) were aligned with PILEUP of a GCG/EMBNET facility. Identical amino acids in three or more out of six sequences are on a black background, similar residues in three or more are on a gray background. Underlined with asterisks (*) is a proposed acyl-CoA synthetase signature motif (Black *et al.*, 1987). The double underline at the C terminus of PfACS1 is the Led domain, the *P. falciparum* sequence proposed to be involved in the extra-parasitic localization of the protein.

AMP-binding site and the ATP-binding pocket within the fatty acyl-CoA synthetase (Conti *et al.*, 1996; Black *et al.*, 1987; Johnson *et al.*, 1994). A highly conserved 25 amino acid residue segment within this region is common to all fatty acyl-CoA synthetases (underlined with asterisks in Figure 4) (Black *et al.*, 1987). These alignments revealed that the PfACS1 has a striking longer carboxy-terminal region of 70 residues corresponding to the Lys-enriched domain (Led) of the parasite protein (double-underlined in Figure 4). Interestingly, an N-terminal hydrophobic 20 amino acid residue sequence, characteristic of membrane proteins, was not found in any of these acyl-CoA synthetases analyzed. As determined by SDS-PAGE, the size of PfACS1 is greater (100 kDa) than that of mammalian acyl-CoA synthetases (ACS) (70-80 kDa).

Expression and purification of PfACS1

In order to determine the enzymatic activity of PfACS1, the whole coding sequence, including a six His tag at its C terminus, was produced and purified from baculovirus/Sf9 insect cells. Attempts to express it in different *Escherichia coli* systems, including those lacking the Fad gene (Knoll *et al.*, 1994), were unsuccessful. Toxicity of the protein and mutations introduced by the bacteria accounted for this failure. Thus, Sf9 cells were transfected with recombinant PfACS1 × 6His or a mock virus (non-recombinant) and its expression analyzed by SDS-PAGE and Western blots (Figure 5). Using the Triton X-100 soluble proteins of the transfected Sf9 cells as starting material, and chromatography on Ni^{2+}-NTA resin,

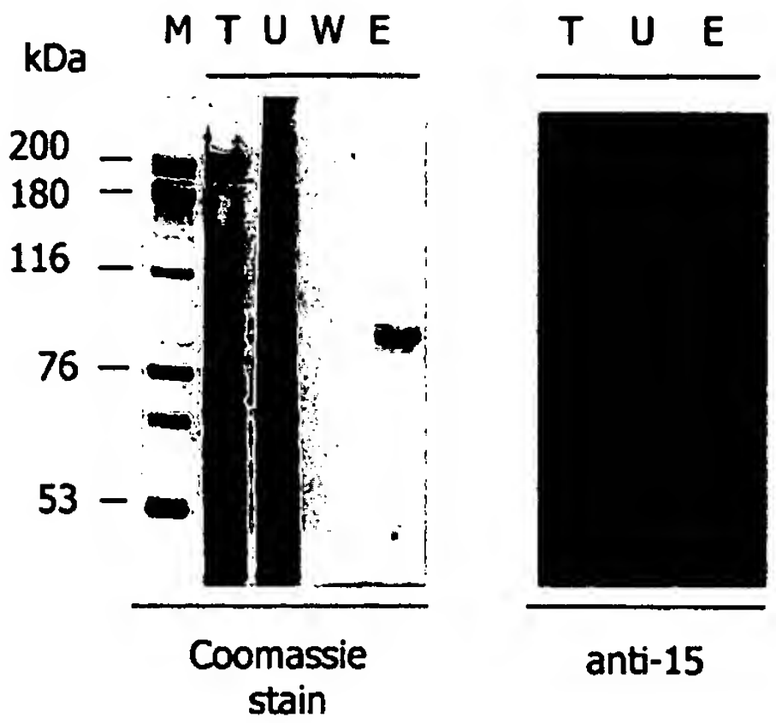

Figure 5. Expression and purification of full-length PfACS1 × 6His in Baculovirus/Sf9 insect cells. PfACS1 × 6His expressed in recombinant Sf9 cells were subjected to chromatography using Ni^{2+}-NTA agarose under native conditions indicated in Materials and Methods. Visualization of protein extracts was performed by Coomassie blue stain of an SDS-7.5% PAGE and Western blots with the anti-15 antibody. M, Molecular mass markers; T, total Sf9 cell extract in the presence of 40 mM imidazole; U, flow-trough unbound extract; W, last-washed extract; E, fraction eluted with 240 mM imidazole, pH 7.5.

PfACS1 × 6His was highly purified (>90%) as shown by SDS-PAGE stained with Coomassie brilliant blue. This major protein of 95 kDa was specifically reactive with the anti-15 antibody in Western blots. Some minor proteins that co-purified with the 100 kDa protein was also reactive with the anti-15 antibody. These minor proteins likely represent degradation products of PfACS1 × 6His.

Long-chain acyl-CoA synthetase activity of the recombinant PfACS1×6His

The ACS activity of the purified recombinant enzyme was determined using [^{3}H]palmitate as a substrate and saturating concentrations of the rest of the co-substrates. As shown in Figure 6, not UTP, AMP, nor GTP at 5 mM may substitute the ATP for optimal activity. AMP, a product of the reaction in AMP-forming enzymes, had an inhibitory effect at concentrations of 5 mM in a reaction mixture containing ATP at 50 µM. These experiments indicated that the purified PfACS1 × 6His was an acyl-CoA synthetase enzyme completely dependent on the presence of ATP, CoA, Mg^{2+}, and fatty acid for activity. The PfACS1 × 6His exhibited an apparent K_m of 19(±3) µM and a specific activity (V_{max}) of 40(±3) nmol/minute per mg of PfACS1 protein at 37°C calculated from a Lineweaver-Burk plot with a substrate concentration ranging from 5 to 50 µM (plot not shown).

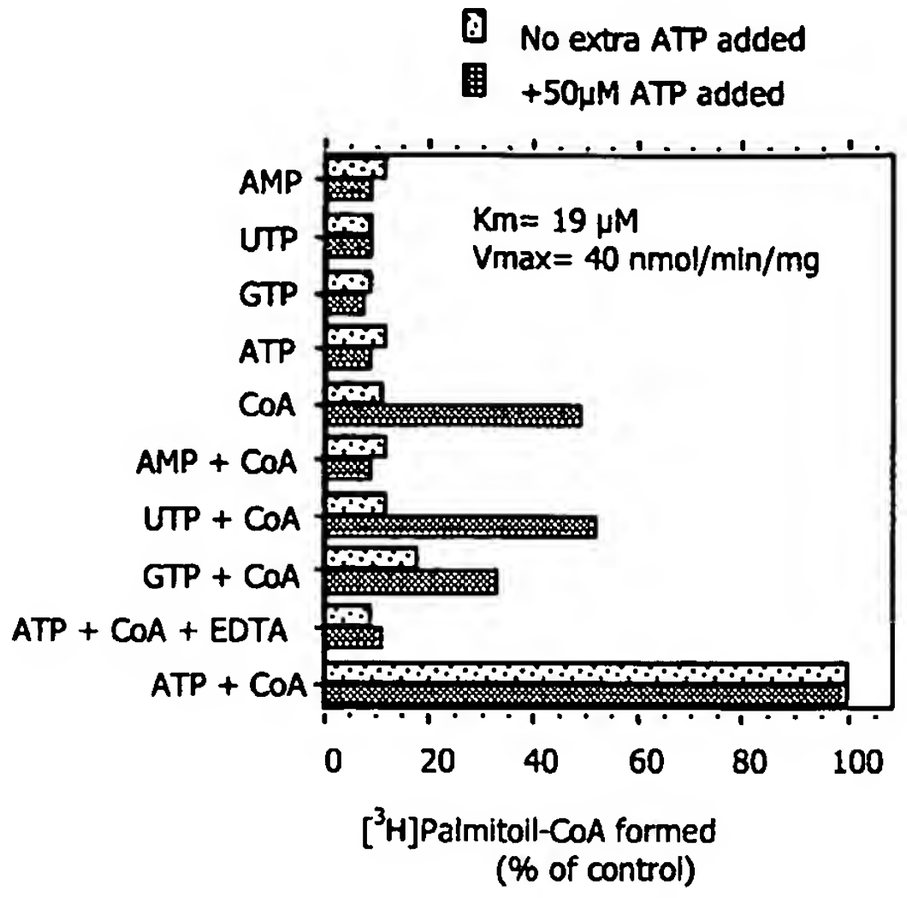

Figure 6. Acyl-CoA synthetase activity of purified recombinant PfACS1 × 6His protein. Effects of co-substrate on the [^{3}H]palmitoyl-CoA synthetase activity of PfACS1 × 6His represented as percentage of activity with respect to the activity obtained in an optimal reaction mixture containing 5 mM ATP, 0.3 mM CoA, 10 mM Mg^{2+} and 0.2 µM [^{3}H]palmitate (1 µCi) for 20 minutes at 37°C, pH 8. In dotted bars, the reaction mixture contained an extra 50 µM ATP. Optimal control reaction (100% activity) = 65,890 cpm. Values are the mean of three assays. The graph shows the V_{max} and K_m of the purified PfACS1 × 6His calculated from a Lineweaver-Burk plot with a substrate concentration ranging from 5 to 50 µM.

The ACS activity of PfACS1 × 6His was optimum at a pH of 8-9 (not shown), a rather basic pH compared with other ACS's organisms.

Immunolocalization of PfACS1 in infected erythrocytes

As shown in the immunofluorescences of Figure 7, asynchronous *P. falciparum*-infected erythrocyte cultures (~10% parasitemia), were positively stained with the polyclonal antibody arose against the C-terminal part of the PfACS1 protein (anti-15 antibody). Neither the pre-immune serum nor the second FITC-conjugated antibody bound to the cells (not shown). The IFA staining revealed the presence of vesicular structures in the erythrocyte cytoplasm (Ec) as early as five hours post-invasion. The staining pattern of infected erythrocytes evolved during the parasite life-cycle in the infected erythrocytes, which lasts for 46-48 hours, likely as a consequence of the accumulation of vesicle structures bearing PfACS1 between the parasite and the parasitophorous vacuolar membrane (pvm). Thus, when the pvm was stained with the antibody, at mid-late stage, many infected

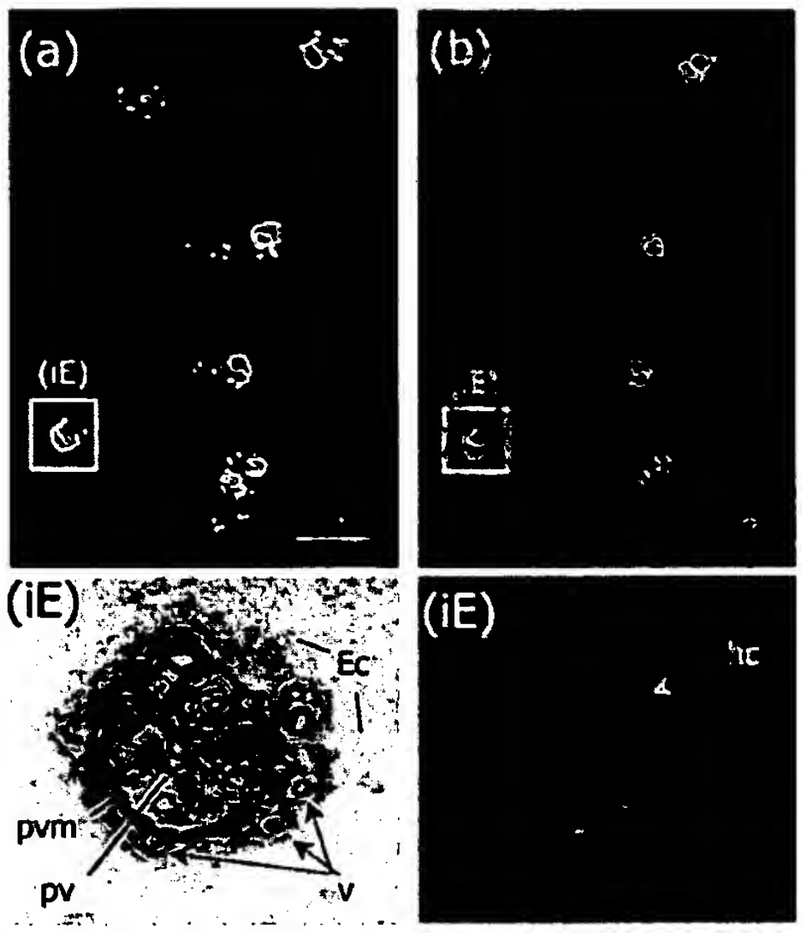

Figure 7. Localization of PfACS1 protein in *P. falciparum*-infected erythrocytes by indirect FITC immunofluorescence microscopy. (a) Asynchronous culture of infected erythrocytes (iE) stained with anti-15 antibody; (b) the same field as (a) using phase-contrast optics. At bottom left is shown a magnified image of the iE selected in (a), and processed with the Adobe Photoshop program to increase contrast and densities and show in relief different parts of this trophozoite-infected erythrocyte, in which Ec stands for erythrocyte cytoplasm; pv, parasitophorous vacuole; pvm, parasitophorous vacuolar membrane; and v, pv budding vesicles. At bottom right is the corresponding magnification of the same iE under phase contrast microscopy, in which the dark hemozoine crystal (hc) is indicated with an arrow. hc is a metabolic product of hemoglobin produced by mature infected erythrocytes (trophozoite and schizonts) in the digestive vacuole of the parasite.

erythrocytes contained cytoplasmic vesicles (v) and membranous protrusions budding off the parasitophorous vacuole (pv) Figure 7, iE, left panel). The presence of a black pigment of hemozoine crystals (hc) in the parasite digestive vacuole is characteristic of mature tropozoite-stage parasite (Figure 7, iE, right panel).

Immunoelectron microscopy of infected erythrocytes labeled with the anti-15 antibody (Figure 8) was in agreement with the immunofluorescence pattern. Although the level of labeling was low, several kinds of vesicle-like structures were shown to contain consistently and specifically PfACS1 epitopes. All these structures were located in the cytoplasm of the infected erythrocyte, outside the parasite. The antibody bound to small irregular structures of ~0.3 μm, in which the epitopes seemed to be on the surface of membranes (micrographs c1) or in linear disposition parallel the erythrocyte plasma membrane, but clearly in the inner place (micrograph B, b1). In addition, the antibody bound to an apparent distinct structure in which the reactivity was localized either in the interior of the vesicles or in protein aggregates. All of these structures may correspond with small Maurer's clefts (100-600 nm diameter) or non-membrane bound aggregates in the host cytoplasm that may be active in the trafficking of proteins and membranes between the parasite and the host

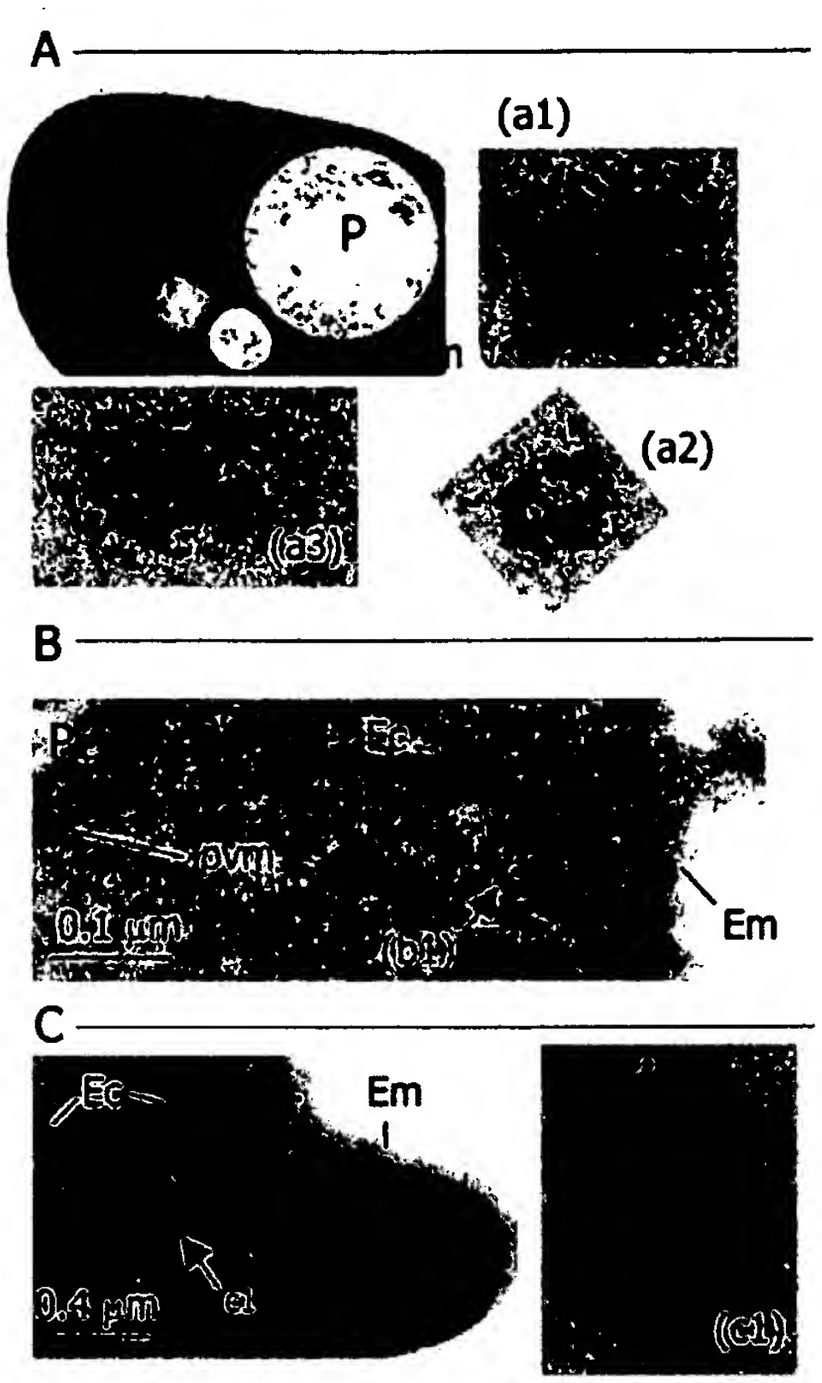

Figure 8. Immunoelectron microscopy of *P. falciparum*-infected erythrocytes labeled with anti-15 antibody and protein G-gold (particles of 10 nm diameter). Arrows in the photomicrogaphs indicate different structures containing PfACS1 protein located in the erythrocyte cytoplasm (Ec) and outside the intracellular parasite (P) in all the cases observed. In structures (a1) and (a3), the PfACS1 appears to be located inside the vesicle or as possible protein aggregates; (a2) is a vesicle-like structure that did not show labeling with the anti-15 antibody; (c1) shows apparent labeling on the surface of membrane structures of this vesicle; and (b1) is a linear string of labeling close of the erythrocyte membrane (Em). Pvm signifies parasitophorous vacuole membrane. All the photographs were taken with a Zeiss electronmicroscope at (a) 10,000× and (b) and (c) 50,000 × . The negative films were scanned-digitalized and the images processed with Adobe Photoshop to get the best contrast.

plasma membrane (Barnwell, 1990; Gormley *et al.*, 1992).

Stage-specificity of the PfACS1 expression

In order to determine what IF pattern corresponded with what asexual stage (Figure 9), samples from highly synchronous cultures were prepared as before and the PfACS1 expression analyzed. After merozoite entry into the erythrocyte, the parasite develops a vacuole (v) and becomes a ring form (R), shaped like a signet ring, with the nucleus as the seal. This stage lasts ~18 hours. The rings grows by increasing its cytoplasm and feeding on the erythrocyte cytoplasm by ingesting hemoglobin, which is digested in food vacuoles. This stage persists for 18-30 hours. During this time, the parasite is the most metabolically active. During the following 16-18 hours, the schizont stage (S), the nucleus divides three to five times to form 8-32 nuclei. At this stage, the parasite occupies most of the erythrocyte and the pigment is concentrated in a single mass. In Figure 9, as early as five to ten hours post-invasion, at the early ring stage, PfACS1 was detected at the erythrocyte cytoplasm in vesicle-like structures. At the mid-trophozoite stage (15-20 hours post-invasion) PfACS1 was detected both at the parasite and spread across the host cytoplasm. From this time on (32 hours and 40 hours), the antibody stained the whole parasite in the limits of the pv and vesicle structures at the erythrocyte cytoplasm (Ec). These observations suggest that the PfACS1 protein is segregated at the space between the parasite and the pvm or the pvm before export into the erythrocyte cytoplasm where it accumulates and makes it possible to be detected with the antibody.

Discussion

We have isolated and characterized a member of the long-chain fatty acid CoA synthetase family from *P. falciparum*-infected erythrocytes as shown by the enzymatic activity of the recombinant product purified from baculovirus-transfected insect cell line. The enzymatic reaction of these type of enzymes is carried out by two-step mechanism that proceeds through the pyrophosphorolysis of ATP (Groot *et al.*, 1976): (1) fatty acid + ATP Mg^{2+} [fatty acid-AMP] + PPi; (2) [fatty acid-AMP]+ CoASH = Fatty acid-S-CoA + AMP. As shown in Figure 6, the enzymatic activity of the recombinant PfACS1 was completely dependent on ATP, Mg^{2+}, CoA and fatty acid, which confirms the activity of the purified protein and the suggestions of the homology sequence analysis. We have shown that the *pf*acs1 gene is transcribed during the erythrocyte stages, and the molecular mass of the denatured protein product (100 kDa) expressed in the parasite is in accordance with the calculated mass of the deduced amino acid sequence. Southern blot analysis revealed several homologous genes in the parasite genome that hybridized with *pf*acs1 gene probes. This situation would be more like what has been found in acyl-CoA synthetases from eukaryotes than in bacteria. Thus, for example, the *S. cerevisiae* acyl-CoA synthetase family comprises at least four members (Faa1p to Faa4p) (Johnson *et al.*, 1994), rat acyl-CoA synthetase is constituted so far by four members (Kang *et al.*, 1997), while only one has been described in *E. coli* (*fadD*) (Black *et al.*, 1987). These data have important biological consequences, since the variability of enzymes brings with it the possibility of different functional roles. In the yeast *S. cerevisiae*, the Faa1p is responsible for activation of

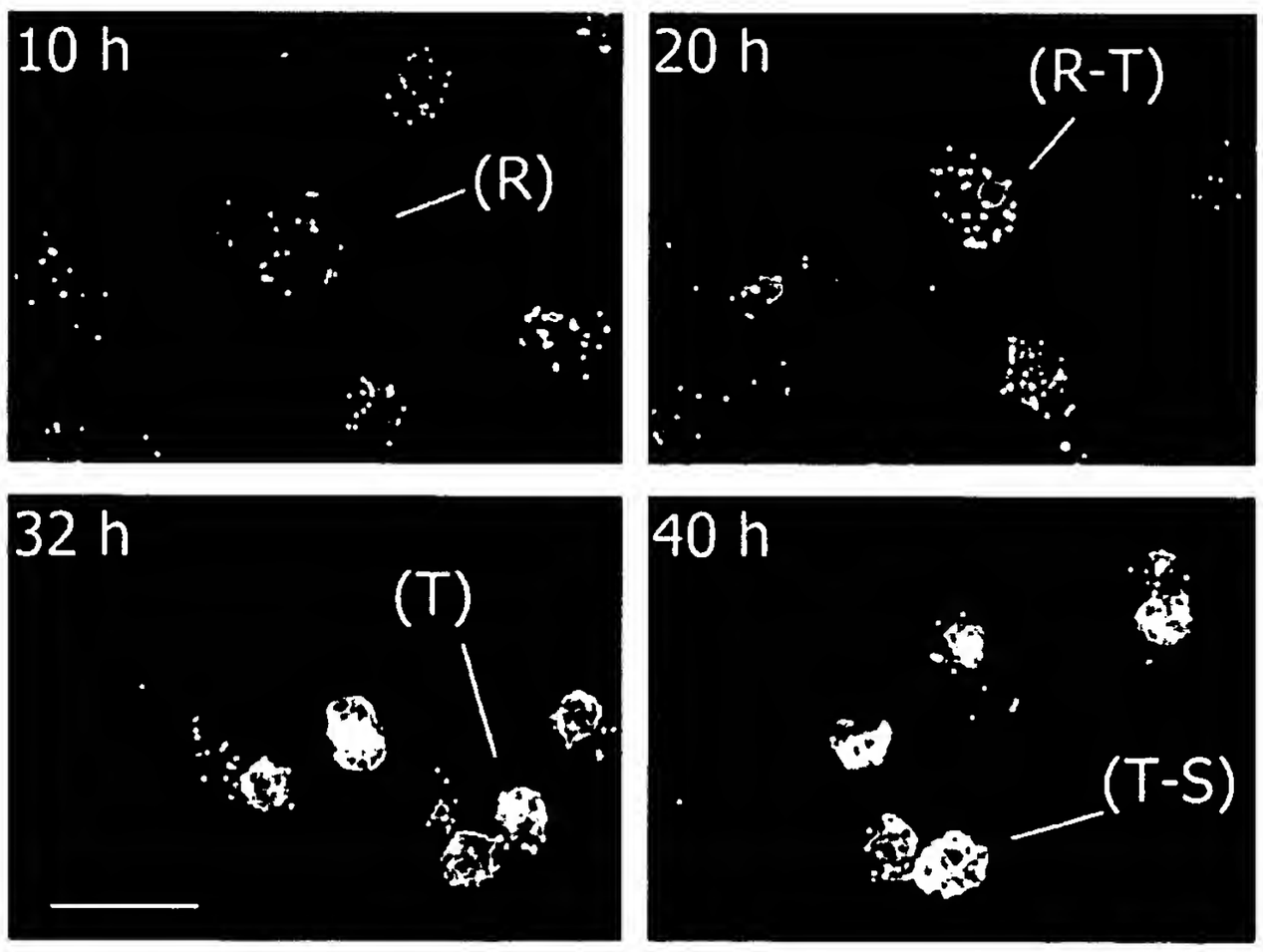

Figure 9. FITC immunofluorescence of stage-specific expression of the PfACS1 protein in synchronized *P. falciparum*-infected erythrocyte with the anti-15 antibody. Cultures were prepared ten hours after synchronization at the ring stage (R); after 20 hours at the ring-trophozoite stage (R-T); after 32 hours at the trophozoite stage (T); and after 40 hours at the trophozoite-schizont stages (T-S). The bar in 32 h represents 10 µm. Immunofloures-cence with the anti-TrpE antibody and normal mouse sera as controls gave a negative staining (not shown).

imported fatty acids to their CoA derivatives, and Faa2p and Faa3p are able to utilize endogenous but not exogenously imported fatty acid substrates (Johnson *et al.*, 1994). Rat acyl-CoA synthetase 4 preferentially uses arachidonate and eicosapentaenoate among C8-C22 saturated fatty acids and C14-C22 unsaturated fatty acids (Kang *et al.*, 1997). In *P. falciparum*, with a complex life-cycle involving different stages in human and insect host cells, it is possible that there exists an expression regulation mechanism that may give rise to stage-specific expression as well as sub-cellular compartmentalization.

The apparent K_m value of the purified recombinant product (PfACS1 × 6His) for palmitate (19 μM) was similar to that observed with enzymes from other sources, as for instance, the yeast *S. cerevisiae* Faa1p × 6His (20 μM) (Knoll *et al.*, 1994). However, the V_{max} of the purified product (40 nmol/minute per mg) was 250 times higher than the V_{max} of *S. cerevisiae* Faa1p × 6His (Knoll *et al.*, 1994). This may be a requirement for the very active lipid metabolism that needs to take place after erythrocyte infection for membrane biogenesis (Haldar *et al.*, 1985; Haldar, 1992; Lauer *et al.*, 1997). While *S. cerevisiae* can activate exogenous fatty acids as well as synthesize them *de novo* from acetate (Schweizer *et al.*, 1978), *Plasmodium* relies only on the exogenous sources (Holz, 1977; Sherman & Greenan, 1984; Haldar *et al.*, 1985; Beaumelle & Vial, 1998).

The immuno-localization of PfACS1 protein in the cytoplasm of the infected erythrocyte, in vesicle-like structures (possibly small Maurer's clefts), together with the fact that the PfACS1 protein remains in the pellet of inside-out-vesicle preparations (data not shown) indicates that it is a membrane-associated protein. This conclusion is further supported by the fact that saponin extracts of infected erythrocytes did not dissolve the PfACS1 protein as assessed by Western blots (Figure 2). This saponin treatment has been shown to fail to separate host erythrocyte membrane from the parasite, regardless of its stage (Beaumelle *et al.*, 1987), suggesting that it could interact with host erythrocyte membrane proteins or with parasite saponin-insoluble proteins (Figure 2). A molecular aspect that further support the interaction of PfACS1 with other proteins is the 70 amino acid residue region located at the C terminus of the PfACS1 protein. This domain, positively charged, is not found in other acyl-CoA synthetases so far described, suggesting that it does not play a role in the enzymatic activity, but rather, it may be involved in extra-parasite interactions. The N-terminal hydrophobic 20 amino acid residue region showed characteristics of a signal sequence (Martoglio & Dobberstein, 1998), and because the cellular localization of PfACS1, it could be required for directing the protein from the parasite toward the erythrocyte cytoplasm. With regard to this, the PfACS1 protein may be of great value to further clarify the processes and the interactions between host and parasite proteins at the molecular level and how the parasite proteins are transported to the erythrocyte membrane, processes that are not completely understood (Knapp *et al.*, 1991; Wiser, 1991; Gormley *et al.*, 1992).

Given the early expression of PfACS1 in infected erythrocytes and the fact that the parasite, in its host erythrocyte, is unable to carry out fatty acid synthesis *de novo* (Holz, 1977; Sherman & Greenan, 1984; Haldar *et al.*, 1985; Beaumelle & Vial, 1998), it is reasonable to assume that this enzyme might play a role in the formation of the tubuvesicular membranes that extends from the parasite's vacuole membrane to the periphery of the red cell. This vesicular network constitutes a transport system that allows efficient access of nutrients to the parasite (Lauer *et al.*, 1997) and it is essential for *Plasmodium* as intracellular parasite. Our data further support the importance of this PfACS1, since its putative role in the activation of fatty acids would take place at the right place where the enzyme and substrates are concentrated (fatty acid from host membrane and serum). From this cellular localization, their fatty acyl-CoA derivatives may be used for phospholipids synthesis and incorporated into the newly synthetasing membranes and fatty acid metabolism. In mammals, this biochemical activity is blocked with Triacsins inhibitors (Tomoda *et al.*, 1991) and with fatty acid analogues in *P. falciparum* (Beaumelle & Vial, 1998). With a view to a possible chemotherapeutic approach, it is interesting that the localization of these kinds of enzymes outside the parasite, in the host erythrocyte cytoplasm, since a potential drug would have to pass through only one bilayer membrane to reach its target protein. The presence of similar enzymes in the human host could make it difficult to find specific inhibitors. However, the high degree of dissimilarity between human acyl-CoA synthetases and *P. falciparum* ACSs (PfACS1) suggests the possibility of finding specific inhibitory drugs with a good therapeutic index.

Material and Methods

Oligonucleotides

Oligonucleotides were synthesized using an automatic DNA synthesizer (Applied Biosystems model 381). The matching sequence at the *Pfacs1* gene as well as the restriction site (underlined) and other traits included in the primer are following indicated: 1165 (nucleotides 1549-1565/GGAGGGCCTATTTGTCC); 1389 (1830-1846/GAATAATCTCTATTCGC); 1393 (1339-1322/CAAGAAAATTAACTAACC); 1859 (1565-1549/GGACAAATAGGCCCACCC); IP8 (2264-2248/CTAATAGGAGCACATCC); IP15 (463-481/AGTGGCGTTACTACAT TAG); IP18 (614-595/TTCTTCAAGTGTGGCAATTC); IP19 with an *Eco*RI site (1-20/GGAATTCTAATGGTATCATTTTAAATAT); IP20 with an *Eco*RI site (2463-2442/GGAATTCTCATTTATTCATTTCTTTTTGA); IP22 (904-923/GAGGATCCTAA TTTGTTAC); IP23 (923-904/GTAACAAAATTAGGATCCT); IP28 with an *Eco*RI site, termination codon (in bold), six histidine codons (second underlined

sequence), and primer (2460-2440/GGGAATTCTTA GTGATGGTGATGGTGATGTTTATTC ATTTCTTTTTG ACG). Matching at the cloning flanking site of λgt10:IP13 with a *Bam*HI site (CGGGATCCGCCT GGTTAAGTCCAAGC); IP14 with a *Bam*HI site (CGGGATCCATGAGTATTTCTTCCAGGGTA).

Parasite culture and cell lines

The 3D7 strain of *P. falciparum* was cultured in human erythrocytes under a low oxygen atmosphere (5% O_2, 5% CO_2 and 90 N_2) according to Trager & Jensen (1976). Synchronization was attained by Percoll gradient centrifugation method (Braun-Breton *et al.*, 1986). Parasites were released from infected erythrocytes by saponin lysis (0.15% (w/v) saponin in PBS) as described (Beaumelle *et al.*, 1987). *Spodoptera frugiperda* Sf9 cells were grown in TC100 medium (Gibco) supplemented with 10% (v/v) fetal calf serum, 100 units/ml penicillin and 100 µg/ml streptomycin.

Libraries, clones and accession numbers

The *Pf*acs1 sequence (Figure 1, scheme) was isolated from a λgt10 *P. falciparum* genomic library that we constructed as described by Anderson & McDonald (1993). Genomic DNA from the *P. falciparum* was digested with *Eco*RI start activity, and the 2-4 kb generated fragments cloned into the *Eco*RI site of λg10 vector (Promega). The first screening was performed using a 715 bp long probe from clone 15 (probe-c, Figure 2) (digoxigenin (DiG)-labeled IP8-1165 fragment, generated by PCR) giving, as a result, clone 34. Clone 15 is a 1222 bp long insert, obtained from a λgt11 mung bean generated genomic library (McCutchan *et al.*, 1984) of the IMMT strain (7G8 clone), which was positive for malaria-falciparum immune sera from different geographical regions (data not shown). The clone 15 sequence has been deposited in the GenBank database with the accession number U10121. The λgt10 clone 34 was PCR-amplified with primers IP13-IP14 and the resulting purified insert was cloned into the pGem-T-vector (Promega) for sequencing. A 151 bp long DIG-labeled probe (probe-a, Figure 2), obtained by PCR with primers IP15 and IP18 from the 5′-end of clone 34, was used for a second screening of the same library giving the clone 27. The whole sequence of *Pf*acs1 has been deposited in the GenBank database under GenBank accession AF007828.

DNA sequence analysis

DNA sequences were determined by PCR/fluorescent nucleotides of both strands using a DNA Sequencing Kit (Perkin Elmer) in an ABI373 DNA sequencer (Applied Biosystems, Foster City, GA) with vector primers as well as the internal ones. The DNA sequence analysis was carried out using Strider 1.1 software (Marck, 1988) and the sequence analysis package Genetics Computer Group version 9.0 (University of Wisconsin, Madison, WI) (Devereux *et al.*, 1984). Searches made by the BLAST program were performed in the non-redundant protein library at the National Center for Biotechnology Information. Comparison analysis was performed with the PILLEUP and BOXSHADE softwares. Prediction of transmembrane structure was carried out at GenomeNet WWW server/MOTIF, Institute for Chemical Research, Kyoto University, Japan.

Southern blot and probes

Genomic DNA was isolated by the SDS-proteinase K method and carried out as indicated by standard methodology (Ausubel *et al.*, 1990). After digestion with various restriction enzymes according to the manufacturer's recommendations, DNA was fractionated on a 1% agarose gel, blotted onto nitrocellulose filters (Schuell and Schuells) and UV cross-linked at 1.2×10^5 µJ/cm^2. Filters were hybridized with [α-^{32}P]ATP-labeled probes obtained by random priming using a DNA labeling kit (Pharmacia Biotech) in 5 × SSC, 15% (w/v) SDS, 1% blocking reagent (Boehringer Mannheim) for 18 hours at 60°C. Blots were washed in 0.2 × SSC, 0.1% (w/v) SDS at 60°C before exposure to Curix RP2 Agfa film at −70°C. The probes-a (primers IP15-IP18, 151 bp long) and b (primers 1389-IP8, 434 bp long) were obtained by PCR amplification from clone 34 DNA, purified, and then radiolabeled.

RNA isolation, Northern blot and probes

Total parasite RNA was extracted using the method described by Chomczynski & Sacchi (1987) and the Northern performed following standard methodology (Ausubel *et al.*, 1990). Briefly, the RNA (10 µg) was separated by 1.2% agarose/formaldehyde gel electrophoresis and transferred onto Biotrans nylon membrane (ICN), UV cross-linked and hybridized for 18 hours at 58°C with random-priming Dig-labeled probe-c (715 bp long, primers 1165-IP8). Blots were washed with 0.5 × SSC, 0.1% SDS at 60°C before applying an immunodetection procedure with chemoluminiscent substrate CDP-star (Boehringer Mannheim) according to the manufacturer's recommendations.

Production of mouse specific antisera against clone 15 (anti-15)

The 1222 bp long insert of the clone 15 coding for an amino acid sequence identical with the carboxy-terminal 397 amino acid residues of the PfACS1 (clone 34) was subcloned into the *Eco*RI site of the pATH11 fusion expression vector (Yansura, 1990) in frame with the anthranilate synthetase (TrpE) bacterial protein. The pATH11 construct was used to transform *E. coli* C600, and induced the expression according the procedures already described (Ausubel *et al.*, 1990; Yansura, 1990). The bacterial extracts containing the fusion protein TrpE-15 (78 kDa) and TrpE (35 kDa) were purified by preparative electro-elution from a SDS-PAGE system as described by Matesanz & Alcina (1996). Female eight-week old Swiss mice were injected with 10 µg of highly purified (>95%, data not shown) fusion protein TrpE-15 or TrpE as control, in three doses over a period of two months: the first dose emulsified in complete Freund's adjuvant, and the second in incomplete Freund's adjuvant, were injected intraperitoneally; the third does was injected intravenously in PBS. The serum was adsorbed of irrelevant cross-reactivity and the specificity confirmed in Western blots of bacterial extracts expressing TrpE-15 fusion protein and TrpE (data not shown).

Western blots

Extracts of *P. falciparum* infected erythrocytes (iE) at 5-10% parasitemia, and uninfected erythrocytes (E), pelleted by centrifugation and lysed with 0.15% saponin

in PBS to release the parasites; and recombinant and mock virus-infected Sf9 cells, and purified recombinant parasite protein (PfACS1 × 6His) were subjected to SDS-PAGE under reducing conditions and Western blots following standard procedures (Ausubel *et al.*, 1990). Blots were incubated with 1/1000 dilution of the mouse anti-15 antibody for four hours at 22 °C, followed by horseradish peroxidase-conjugated secondary antibody incubation (1/15,000 dilution) (Pierce), and developed with the chemoluminiscent substrate ECL (Amershan).

Construction of PfACS1 × 6His expression vector

The whole coding sequence of PFACS1 × 6His (2493 bp long) containing an added tag of 6 × His at the C-terminal was obtained by PCR reaction using Pfu *Taq* polymerase (Stratagene) with 10 ng of *P. falciparum* DNA and the primers IP19-IP28 in the following reaction conditions: a single five minute, 95 °C denaturation step, 35 cycles at 95 °C for 0.75 minute; 47 °C for 1.5 minutes; 72 °C, three minutes; and ten minutes at 72 °C as a final extension step. The PCR-amplified insert was cloned into the *Eco*RI site of the pFastBac1 plasmid of the Bac-to-Bac Baculovirus Expression System (Life Technologies, Gibco-BRL) and transfected into *E. coli* DH10Bac as indicated by the manufacturer. The sequence was then confirmed. The resulting recombinant bacmid was transfected by lipofection into Sf9 cells and the supernatant containing recombinant or mock virus collected for following infections. The PfACS1 × 6His expression by the Sf9 infected cells were monitored by Western blots with anti-15 antibody.

Purification of PfACS1× 6His from recombinant Sf9 cells

All purification procedures were carried out at 4 °C. Twenty dishes (100 mm diameter) with 7×10^6 recombinant or mock infected Sf9 cells were harvested at 48 hours after infection, and resuspended in 6 ml of lysis buffer A (50 mM potassium phosphate (pH 7.4), 300 mM potassium chloride, 400 mM sodium chloride, 40 mM imidazole, 20 % (w/v) glycerol, 1 % (w/v) Triton X-100, 4 mM 2-mercaptoethanol and 1 mM phenylmethylsulfonyl fluoride (PMSF)), disrupted by sonication and centrifuged at 10,000 *g* for ten minutes. The resulting supernatant was added to 300 µl of buffer A-equilibrated Ni^{+2}-NTA resin (QIAGEN Inc.), and the slurry was stirred for four hours at 4 °C. The resin was washed four times with five volumes of buffer A, and three further washes with buffer A containing 2 mM ATP. Protein was eluted with 300 µl of buffer A containing 2 mM ATP, and 240 mM imidazole. Protein concentration of eluted active fractions was measured with a Bicinchonic acid/$CuSO_4$ protein determination kit (Sigma) and by densitometry of the band in Coomassie blue-stained gel compared with stained molecular mass standard (Pharmacia Biotec). It was between 8 and 12 µg/ml. Extracts of mock-infected Sf9 cells were subjected to the same procedure.

Enzymatic activity of purified PfACS1×6His

Acyl-CoA synthetase activity was determined by the isotopic assay described by Wilson *et al.* (1982) using [9,10(n)-^{3}H] palmitic acid in ethanol (52 Ci/mmol) (Amersham Corp., Buckinghamshire, United Kingdom) as one of the co-substrate. Briefly, 100 µl of incubation mixture contained 2 µl of purified preparation of PfACS1 × 6His (20 ng) in Tris-HCl buffer (pH 8) (100 mM Tris-HCl, 10 mM MgCl2, 1 mM KF, 2 mM EDTA, 2 mM dithiothreiol), 5 mM ATP, 300 µM CoA and 0.2 µM [^{3}H]palmitic acid (1 µCi), was incubated for 20 minutes at 37 °C. The reaction was stopped by the addition of 125 µl of Dole's solution (isopropyl alcohol, heptane, 1 M H_2SO_4, 40:10:1, by vol.) and 50 µl of water followed by 500 µl of heptane and vigorous vortexing. After centrifugation for one minute at 10,000 *g* the upper phase was discarded. The lower aqueous phase was then washed twice more: 500 µl of scintillation liquid was added to each tube and counted in a MicroBeta Plus counter (Wallac, EG&G Company, Turku, Finland). The K_m and V_{max} of the PfACS1 shown inside the plot of Figure 6 was determined in a reaction mixture containing palmitic acid from 5 to 50 µM by applying the Lineweaver-Burke plot (not shown). The effect of pH was carried out in a reaction mixture containing Mops as the buffer (Knoll *et al.*, 1994). The purified fraction on Ni^{2+}-NTA resin from mock-infected Sf9 cells gave no activity.

Indirect immunofluorescence assay (IFA)

Thin smears of parasitized erythrocytes from asynchronous or synchronous cultures of *P. falciparum* were air-dried and fixed in cold acetone for ten minutes and processed for IFA as described (Alcina *et al.*, 1986). Fixed preparations were overlaid with 25 % (v/v) fetal calf serum (FCS) in PBS for 15 minutes, and incubated with the first antisera (anti-15 or anti-TrpE) at 1/500 dilution for two hours. A second and third reagent, (anti-mouse biotinylated antibody and fluorescein isothiocyanate (FITC)-conjugated streptavidin, respectively) were used at dilutions as recommended by the manufacturer (Boehringer Mannheim). Photomicrographs were taken using a Carl Zeiss Axiophot microscope.

Immunoelectron microscopy

Localization of the PfACS1 antigen in infected erythrocytes by immunogold labeling was performed following a modified procedure as described (Hernández-Munaín *et al.*, 1991). Briefly, infected erythrocyte cultures at about 8 % parasitemia were fixed in 3 % (v/v) glutaraldehyde/25 mM sodium cacodylate buffer (pH 6.9) for 14 hours. After washing in cacodylate buffer, cells were pelleted into agarose. The pellets were dehydrated in ethanol and finally in propylene oxide. Agarose blocks containing the cells were embedded in Epon 812 resin and polymerized at 60 °C for two days. Ultramicrotome sections of 70 nm were picked up on 200-mesh carbon-coated gold grids. The grids were then placed in a moist chamber for the following incubations: one hour in 25 % FCS/PBS, followed by two hours incubation with 1/500 dilution of the anti-15 antibody in FCS/PBS. Negative controls included anti-TrpE antibody or no first antibody. The second reactive was a gold-labeled protein G of particle size 10 nm (Sigma). The grids were washed in PBS and water and finally they were post-stained with 2 % (w/v) uranyl acetate.

Acknowledgments

We thank Dra Concha Delgado from the Blood Bank in Granada, Spain, for supplying human serum and

erythrocytes and Dr Virgilio Do Rosario from the IHMT Lisbon, Portugal, for the 3D7 strain of *P. falciparum*.

This work was supported by grants from CICYT (PN Bio 93-736 and SAF 97-0043).

References

Alcina, A., Hargreaves, A. J., Avila, J. & Fresno, M. (1986). A *Trypanosoma cruzi* monoclonal antibody that recognizes a superficial tubulin-like antigen. *Biochem. Biophys. Res. Commun.* **139**, 1176-1183.

Anderson, B. & McDonald, G. (1993). Construction of DNA libraries of A-T rich organisms using EcoRI start activity. *Anal. Biochem.* **211**, 325-327.

Ausubel, F. A. R., Brent, R. E., Kingston, D. D., Moore, J. G., Seidman, J. A., Smith, J. A. & Struhl, K. (1990). *Current Protocols in Molecular Biology*, Wiley Interscience, New York.

Barnwell, J. W. (1990). Vesicle mediated transport of membrane and proteins in malaria infected red blood cells. *Blood Cells*, **16**, 379-395.

Beaumelle, B. D. & Vial, H. J. (1988). Acyl-CoA synthetase activity in *Plasmodium knowlesi*-infected erythrocytes displays peculiar substrate specificities. *Biochim. Biophys. Acta*, **958**, 1-9.

Beaumelle, B. D. & Vial, H. J. (1998). Correlation of the efficiency of fatty acid derivatives in suppressing *Plasmodium falciparum* growth in culture with their inhibitory effect on acyl-CoA synthetase activity. *Mol. Biochem. Parasitol.* **28**, 39-42.

Beaumelle, B. D., Vial, H. J. & Philippot, J. R. (1987). Reevaluation, using marker enzymes, of the ability of saponin and ammonium chloride to free *Plasmodium* from infected erythrocytes. *J. Parasitol.* **73**, 743-748.

Black, P. N., Zhang, Q., Weimar, J. D. & DiRusso, C. C. (1987). Mutational analysis of a fatty acyl-Coenzyme A synthetase signature motif identifies seven amino acid residues that modulate fatty acid substrate specificity. *J. Biol. Chem.* **272**, 4896-4903.

Braun-Breton, C., Jendoubi, M., Brunet, E., Perrin, L., Scaife, J. & Pereira da Silva, L. H. (1986). In vivo time course of synthesis and processing of major schizont membrane polypeptides in *Plamodium falciparum*. *Mol. Biochem. Parasitol.* **20**, 33-43.

Conti, E., Franks, N. P. & Brick, P. (1996). Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure*, **4**, 287-298.

Chomczynski, P. & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387-395.

Glick, B. S. & Rothman, J. E. (1987). Possible role for fatty acyl-coenzyme A in intracellular protein transport. *Nature*, **326**, 309-312.

Gormley, J. A., Howard, R. J. & Tarashi, T. F. (1992). Trafficking of malarial proteins to the host cell cytoplasm and erythrocyte surface membrane involves multiple pathways. *J. Cell Biol.* **119**, 1481-1495.

Groot, P. H. E., Scholte, H. R. & Hulsmann, W. C. (1976). Fatty acid activation: specificity, localization, and function. *Advan. Lipid Res.* **14**, 75-126.

Haldar, K. (1992). Lipid transport in *Plasmodium*. *Infect. Ag. Dis.* **1**, 254-262.

Haldar, K., Ferguson, M. A. J. & Cross, G. A. M. (1985). Acylation of a *Plasmodium falciparum* merozoite surface antigen via sn-1,2-diacyl glycerol. *J. Biol. Chem.* **260**, 4969-4974.

Hernández-Munaín, C., Fernández, M. A., Alcina, A. & Fresno, M. (1991). Characterization of a glycosylphosphatidylinositol-anchored membrane protein from *Trypanosoma cruzi*. *Infect. Immun.* **59**, 1409-1416.

Holz, G. G., Jr (1977). Lipids and the malarial parasite. *Bull. World Health Organ.* **55**, 237-248.

Hopp, T. P. & Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl Acad. Sci. USA*, **78**, 3824-3828.

Johnson, D. R., Knoll, L. J., Levin, D. E. & Gordon, J. I. (1994). *Saccharomices cerevisiae* contains four fatty acid activation (FAA) genes: an assessment of the role in regulating protein N-myristoylation and cellular lipid metabolism. *J. Cell Biol.* **127**, 751-762.

Kang, M. J., Fujino, T., Sasano, H., Minekura, H., Yabuki, N., Nagura, H., Lijima, H. & Yamamoto, T. T. (1997). A novel arachidonate-preferring acyl-CoA synthetase is present in steroidogenic cells of the rat adrenal, ovary, and testis. *Proc. Natl Acad. Sci. USA*, **94**, 2880-2884.

Knapp, B., Hundt, E. & Lingelbach, K. R. (1991). Structure and possible function of *Plasmodium falciparum* proteins exported to the erythrocyte membrane. *Parasitol. Res.* **77**, 277-282.

Knoll, L. K., Russell, D. & Gordon, J. I. (1994). Biochemical studies of three *Saccharomyces cerevisiae* acyl-CoA synthetases, Faa1p, Faa2p, and Faa3p. *J. Biol. Chem.* **269**, 16348-16356.

Korchak, H. M., Kane, L. H., Rossi, M. W. & Corkey, B. E. (1994). Long chain acyl coenzyme A and signaling in neutrophils. An inhibitor of acyl coenzyme A synthetase, triacsin C, inhibits superoxide anion generation and degranuolation by human neutrophils. *J. Biol. Chem.* **269**, 30281-30287.

Lauer, S. A., Rathod, P. K., Ghori, N. & Haldar, K. (1997). A membrane network for nutrient import in red cells infected with the malaria parasite. *Science*, **276**, 1122-1124.

Marck, C. (1988). 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucl. Acids Res.* **1**, 1829-1836.

Martoglio, B. & Dobberstein, B. (1998). Signal sequences: more than just greasy peptides. *Trends Cell Biol.* **8**, 410-415.

Matesanz, F. & Alcina, A. (1996). Glutamine and tetrapeptide repeat variations affect the biological activity of different mouse interleukin-2 alleles. *Eur. J. Immunol.* **26**, 1675-1682.

McCutchan, T. F., Hansen, J. L., Dame, J. B. & Mullins, J. A. (1984). Mung bean nuclease cleaves *Plasmodium* genomic DNA at sites before and after genes. *Science*, **225**, 625-628.

McLaughlin, S. & Aderem, A. (1995). The myristoyl-electrostatic switch: a modulator of reversible protein-membrane interactions. *Trends Biochem. Sci.* **20**, 270-276.

Rathod, P. K., McErlean, T. & Lee, P. (1997). Variations in frequencies of drug resistance in *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA*, **94**, 9389-9393.

Schweizer, E., Werkmeister, K. & Jain, M. K. (1978). Fatty acid biosynthesis in yeast. *Mol. Cell. Biochem.* **21**, 95-107.

Sherman, I. W. & Greenan, J. R. T. (1984). Altered red cell membrane fluidity during schizogonic development of malarial parasites (*Plasmodium falciparum* and *P. lophurae*). *Trans. Roy Soc. Trop. Med. Hyg.* **78**, 641-644.

Simoes, A. P., Roelofsen, B. & Op den Kamp, J. A. F. (1992). Incorporation of free fatty acids can explain alterations in the molecular species of composition of phosphatidylcholine and phosphatidylethanolamine in human erythrocytes as induced by *Plasmodium falciparum*. *Cell Biol. Int. Rep.* **16**, 533-545.

Sturchler, D. (1989). How much malaria is there worldwide? *Parasitol. Today*, **5**, 39-40.

Tomoda, H., Igarashi, K., Cyong, J. C. & Omura, S. (1991). Evidence for and essential role of long chain acyl-CoA synthetase in animal cell proliferation. *J. Biol. Chem.* **266**, 4214-4219.

Trager, W. & Jensen, J. B. (1976). Human malaria parasites in continuous culture. *Science*, **193**, 673-675.

Wilson, D. B., Prescott, S. M. & Majerus, P. W. (1982). Discovery of an arachidonoyl coenzyme A synthetase in human platelets. *J. Biol. Chem.* **257**, 3510-3515.

Wiser, M. F. (1991). Malarial proteins that interact with the erythrocyte membrane and cytoskeleton. *Exp. Parasitol.* **73**, 515-523.

Yansura, D. G. (1990). Expression as trpE fusion. *Methods Enzymol.* **185**, 161-166.

Edited by M. Yaniv

(Received 6 April 1999; received in revised form 5 June 1999; accepted 25 June 1999)

JOURNAL OF BACTERIOLOGY, Feb. 1996, p. 871–880
0021-9193/96/$04.00+0

Vol. 178, No. 3

Molecular Analysis of the Anaerobic Succinate Degradation Pathway in *Clostridium kluyveri*

BRIGITTE SÖHLING† AND GERHARD GOTTSCHALK*

Institut für Mikrobiologie, Georg-August-Universität Göttingen, D-37077 Göttingen, Germany

Received 14 July 1995/Accepted 30 November 1995

A region of genomic DNA from *Clostridium kluyveri* was cloned in *Escherichia coli* by a screening strategy which was based on heterologous expression of the clostridial 4-hydroxybutyrate dehydrogenase gene. The gene region (6,575 bp) contained several open reading frames which encoded the coenzyme A (CoA)- and $NADP^+$-dependent succinate-semialdehyde dehydrogenase (*sucD*), the 4-hydroxybutyrate dehydrogenase (*4hbD*), and a succinyl-CoA:CoA transferase (*cat1*), as analyzed by heterologous expression in *E. coli*. An open reading frame encoding a putative membrane protein (*orfY*) and the 5′ region of a gene encoding a σ^{54}-homologous sigma factor (*sigL*) were identified as well. Transcription was investigated by Northern (RNA) blot analysis. Protein sequence comparisons of SucD and 4HbD revealed similarities to the *adhE* (*aad*) gene products from *E. coli* and *Clostridium acetobutylicum* and to enzymes of the novel class (III) of alcohol dehydrogenases. A comparison of CoA-dependent aldehyde dehydrogenases is presented.

The gram-positive anaerobic bacterium *Clostridium kluyveri* (3) ferments ethanol and acetate to butyrate, caproate, and molecular hydrogen (7). ATP, required for growth, is gained by substrate-level phosphorylation from acetyl phosphate, and the quantity is proportional to the amount of hydrogen produced (46, 50). Investigations of additional metabolic abilities revealed that this organism can utilize crotonate, vinylacetate, and 4-hydroxybutyrate as substrates (4, 5) and is able to ferment the unusual substrate combination of succinate plus ethanol (27). A pathway was proposed, one in which succinate is first activated and then reduced by a two-step reaction to give 4-hydroxybutyrate, which is then further metabolized to crotonyl-coenzyme A (CoA) (Fig. 1) (27). In a previous study, we discussed enzymes involved in the anaerobic breakdown of succinate by *C. kluyveri*, specifically, a succinyl-CoA:CoA transferase, a CoA- and $NADP^+$-dependent succinate-semialdehyde dehydrogenase, and a 4-hydroxybutyrate dehydrogenase (49). Wolff et al. (55) independently confirmed these data by ^{13}C-nuclear magnetic resonance studies as well as enzymatic investigations on the dehydrogenases. 4-Hydroxybutyryl-CoA dehydratase, which catalyzes the last step of the succinate-specific pathway, the dehydration and isomerization of 4-hydroxybutyryl-CoA to crotonyl-CoA, was recently identified, purified, and characterized (45). We present here some molecular aspects of this pathway, including the cloning, sequencing, and heterologous expression of a *C. kluyveri* DNA region which encodes a succinyl-CoA:CoA transferase, the succinate-semialdehyde dehydrogenase, and the 4-hydroxybutyrate dehydrogenase.

MATERIALS AND METHODS

Bacterial strains, plasmids, media, and growth conditions. *C. kluyveri* (DSM 555) was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany. *Escherichia coli* JM109 (58) and the pBluescript SK vector (Stratagene, San Diego, Calif.) were from the laboratory collection. *C. kluyveri* cells used for DNA isolation were grown at 37°C under strictly anaerobic conditions on ethanol (300 mM) and succinate (100 mM) as previously described (49). Cells for RNA preparation were cultured in the same medium except that succinate was replaced by acetate (100 mM). *E. coli* cultures were routinely grown at 30°C in Luria-Bertani (LB) medium (42) on a rotary shaker. Tetrazolium indicator plates (6) containing 4 g of 4-hydroxybutyrate per liter were employed for the screening procedure (oxidation of 4-hydroxybutyrate). Utilization of 4-hydroxybutyrate (4 g/liter) as a carbon source for recombinant *E. coli* clones was investigated in M9 medium (42), supplemented with a small amount of yeast extract (0.2 g/liter), $MgSO_4$ (2 mM), and $CaCl_2$ (0.1 mM). Ampicillin (75 mg/liter) was added to the media for *E. coli* as a selection marker, when needed.

Nucleic acids isolation and recombinant DNA techniques. Chromosomal DNA from *C. kluyveri* was isolated by the method of Saito and Miura (41). Total RNA from *C. kluyveri* was isolated by the hot phenol-chloroform procedure as described by Gerischer and Dürre (19). DNA was manipulated by standard methods (42); restriction enzymes and T4 DNA ligase were purchased from GIBCO/BRL (Eggenstein, Germany). For plasmid isolation from *E. coli*, the Quiagen Midi Kit (Diagen GmbH, Düsseldorf, Germany) was used. For cloning purposes, genomic *C. kluyveri* DNA was partially *Hin*dIII digested and fractionated on a sucrose gradient (10 to 40% [wt/vol]). Fractions of approximately 3 to 5 and 4 to 7 kb were ligated into *Hin*dIII-digested pBluescript SK vector, and the product was used to transform *E. coli* JM109. Recombinant clones were screened for their ability to oxidize 4-hydroxybutyrate (see above). Nested deletion subclones were prepared from pCK1 and pCK3 (the clostridial inserts were in a different orientation within the vector) by using the *Kpn*I and *Cla*I sites of the vector to generate the exonuclease III-resistant and -sensitive ends, respectively. Exonuclease III digestion and all further steps were performed with the Erase-a-Base system (Promega, Madison, Wis.) according to the manufacturer's instructions.

DNA sequencing and sequence analysis. Double-stranded DNA was sequenced by the dideoxy chain termination method (43) with α-^{35}S-dATP (DuPont, NEN Research Products) and the Sequenase T7 DNA polymerase kit from U.S. Biochemical (Bad Homburg, Germany) according to the corresponding protocol. The entire sequence (6,575 nucleotides) of the *C. kluyveri* insert of pCK1 was determined for both strands with the nested deletion subclones generated from pCK1 and pCK3 and the commercially available M13/pUC universal and reversal sequencing primers. In addition, some synthetic oligonucleotides (17-mers) complementary to the already sequenced templates were employed. These primers were prepared with a Gene Assembler Plus (Pharmacia Biotech Europe, Freiburg, Germany) according to the manufacturer's instructions.

Computer sequence analysis. The DNA sequence data and the deduced amino acid sequences were analyzed with the Genetics Computer Group Inc. sequence analysis software package, version 6.2 (13), on a VAX 9000 computer. Database searches were performed with the National Biomedical Research Foundation Protein Information Resource Network Server according to the algorithm of Pearson and Lipman (38).

Hybridization. Total chromosomal DNA from *C. kluyveri* or *E. coli* plasmid DNA was digested to completion with the appropriate restriction enzyme and separated on agarose gels. Southern blots on nylon membranes (GeneScreen Plus; DuPont, NEN Research Products) were prepared according to the manufacturer's protocol. DNA fragments used as probes were isolated from agarose gels with the Gene Clean Kit (Bio 101, La Jolla, Calif.). The probes were labeled

* Corresponding author. Mailing address: Institut für Mikrobiologie, Georg-August-Universität, Grisebachstrasse 8, D-37077 Göttingen, Germany. Phone: 49-551-393781. Fax: 49-551-393793.

† Present address: Institut für Mikrobiologie, Martin-Luther-Universität Halle-Wittenberg, D-06099 Halle, Germany.

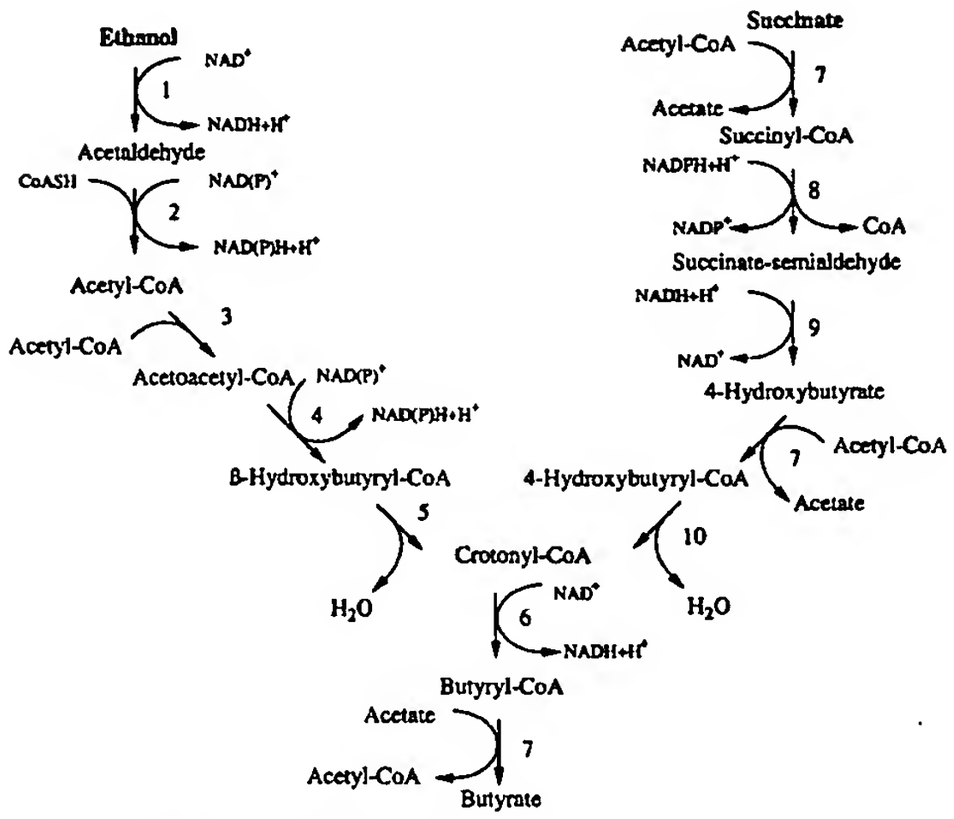

FIG. 1. Schematic pathway for the ethanol-succinate fermentation of *C. kluyveri*. 1, alcohol dehydrogenase; 2, acetaldehyde dehydrogenase; 3, thiolase; 4, β-hydroxybutyryl-CoA dehydrogenase; 5, crotonase; 6, butyryl-CoA dehydrogenase; 7, CoA-transferase (probably several enzymes with different substrate specificities); 8, succinate-semialdehyde dehydrogenase; 9, 4-hydroxybutyrate dehydrogenase; 10, 4-hydroxybutyryl-CoA dehydratase-vinylacetyl-CoA Δ^3-Δ^2-isomerase, as described elsewhere (21, 45, 49, 55). The formation of acetate, hydrogen, and caproate is not shown.

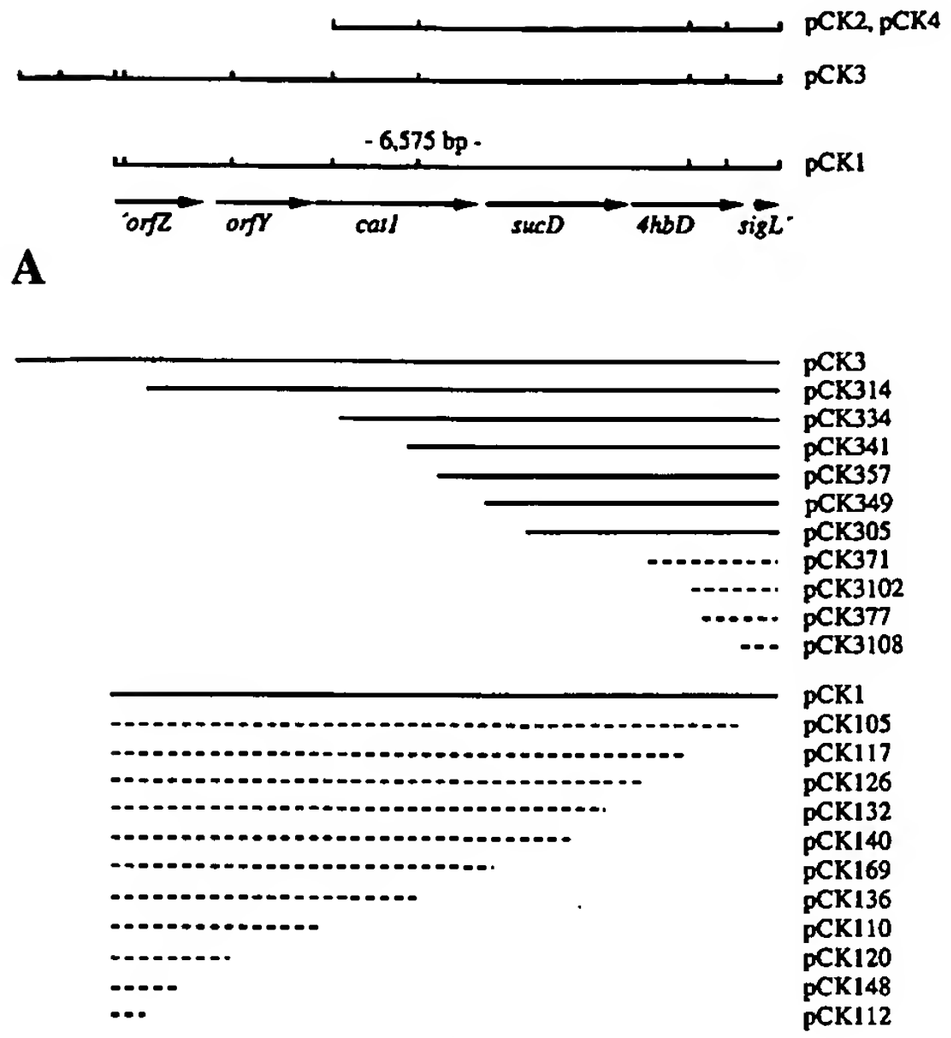

FIG. 2. (A) Schematic map of the pCK1 to pCK4 inserts and the ORFs deduced from the pCK1 nucleotide sequence. Vertical bars indicate *Hin*dIII restriction sites. (B) Schematic drawing of some of the subclones generated from pCK1 and pCK3 by the nested deletion method. The *E. coli* host phenotype (oxidation of 4-hydroxybutyrate) is indicated by continuous lines (positive) and dashed lines (negative).

either with [α-^{32}P]dATP (DuPont, NEN Research Products) by using the Random Primed DNA Labeling Kit (U.S. Biochemical) or with digoxigenin by using the DIG DNA Labeling Kit (Boehringer GmbH, Mannheim, Germany). Membranes were prehybridized in 0.15% (wt/vol) polyvinylpyrrolidone–0.15% (wt/vol) bovine albumin–0.15% (wt/vol) Ficoll 400–0.9 M NaCl–10% (wt/vol) dextran sulfate–1% (wt/vol) sodium dodecyl sulfate (SDS)–6 mM EDTA–90 mM Tris-HCl (pH 7.5)–100 μg of herring sperm DNA per ml (80 ml/cm^2) for 1 to 3 h at 55°C. The appropriate probe was added to the prehybridization solution and incubated for 10 to 15 h at 55°C. Membranes were then washed twice in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) for 15 min at room temperature and once in 2× SSC plus 1% (wt/vol) SDS at hybridization temperature. If necessary, additional washing was performed in SSC solutions, containing 1% (wt/vol) SDS, of decreasing ionic strength (1× SSC, 0.1× SSC) at hybridization temperature. Membranes were then subjected to autoradiography or, in the case of the digoxigenin-labeled probes, manipulated with the appropriate detection kit (Boehringer Mannheim GmbH). RNA for Northern blot analysis was separated in denaturing formaldehyde gels and transferred to nylon membranes (GeneScreen Plus; DuPont, NEN Research Products) as described in the manufacturer's manual. An RNA ladder (0.24, 1.4, 2.4, 4.4, 7.5, and 9.5 kb; GIBCO/BRL) was included as standard for size determination. Hybridization and washing were performed according to the protocol described above for the radiolabeled DNA hybridization procedure.

Determination of enzyme activity. Cells from recombinant *E. coli* clones (pCK1, pCK2, pCK3, pCK4, and pSK) were grown at 30°C in LB medium on a rotary shaker and harvested by centrifugation. If necessary, isopropyl-β-D-thiogalactopyranoside (IPTG; 1 mM) was added at an optical density of 0.5 (580 nm) and cells were grown for an additional 3 h and harvested as described above. In order to prevent enzyme inactivation by oxygen, crude extracts were prepared anaerobically in 50 mM potassium phosphate buffer (pH 7.5) containing 3 mM dithioerythritol, with a French pressure cell (80 MPa; Amicon, Silver Springs, Fla.). Protein concentration was determined according to the method of Bradford (8). 4-Hydroxybutyrate dehydrogenase was assayed in anaerobic glass cuvettes at 30°C in 90 mM 2-amino-2-methyl-1,3-propanediol–HCl (pH 8.5), containing 3 mM dithioerythritol, 1 mM $MgSO_4$, and 1 mM NAD^+ in a final volume of 1 ml. The reaction was started by the addition of 4-hydroxybutyrate (10 mM) and monitored at 340 nm (ε = 6.3 mM^{-1} cm^{-1}). Activity of succinate-semialdehyde dehydrogenase was assayed in anaerobic glass cuvettes in 50 mM TAPS (*N*-tris[hydroxymethyl]methyl-3-aminopropanesulfonic acid)-HCl buffer (pH 8.5) containing 1 mM dithioerythritol–1 mM $NADP^+$ at 30°C in a final volume of 1 ml. The reduction of $NADP^+$ was monitored at 340 nm. To determine the CoA-independent enzyme activity, the reaction was initiated by the addition of succinate-semialdehyde (10 mM). In the case of the CoA-dependent enzyme activity, the assay was initiated with succinate-semialdehyde (10 mM) and then started with CoA (0.1 mM). CoA-dependent succinate-semialdehyde dehydrogenase activity was defined as the difference between the CoA-dependent and CoA-independent enzyme activity. Succinyl-CoA:CoA transferase was determined with succinyl-CoA (0.1 mM) and acetate (0.2 M) as substrates as described by Scherf and Buckel (44). The reaction was initiated with succinyl-CoA, and the acetyl-CoA product was condensed with oxaloacetate-liberating CoASH. Formation of the latter was determined with 5,5'-dithio-bis(2-nitrobenzoic acid) (NbS_2) at 412 nm (ε = 13.6 mM^{-1} cm^{-1}).

Determination of the N-terminal amino acid sequence. The N-terminal amino acid sequence of the purified succinate-semialdehyde dehydrogenase was determined directly from the electroblotted polyvinylidene difluoride membrane by using a Protein Peptide Sequencer 477A (Applied Biosystems, Foster City, Calif.). Detection was performed on-line with a phenylthiohydantoin analyzer.

Nucleotide sequence accession number. The sequence data reported here were submitted to the EMBL database and assigned accession no. L21902.

RESULTS

Cloning of the *C. kluyveri* gene region. Of 10^4 transformants screened, four *E. coli* clones formed red colonies; they were able to reduce the tetrazolium salt added at the expense of 4-hydroxybutyrate oxidation. pCK1, which was chosen for further analysis, harbored, as shown by sequence analysis, a 6,575-bp insert, which contained seven *Hin*dIII restriction endonuclease recognition sites (Fig. 2). pCK2 and pCK4 covered only a part of pCK1 (four *Hin*dIII sites; 4,458 bp) whereas the insert of pCK3 (about 7,500 bp) exceeded the pCK1 region at the left-hand end (Fig. 2). The inserts of pCK1 and pCK2 were in the orientation of the pBluescript *lac* promoter, while those from pCK3 and pCK4 were in opposite orientation. All clones containing the recombinant plasmids pCK1, pCK2, pCK3, and pCK4 were able to grow on 4-hydroxybutyrate as the sole carbon and energy source, which is not a substrate for the *E. coli* wild type. 4-Hydroxybutyrate dehydrogenase activity was detected in all four clones (50 to 160 mU/mg).

To confirm an identical arrangement of the *Hin*dIII fragments within the genomic DNA of *C. kluyveri*, Southern blot analysis was performed. With a 1,087-bp *Eco*RI and an 864-bp *Hin*dIII fragment (bp 900 to 1987 and 2117 to 2981, respectively) as probes for chromosomal *C. kluyveri* DNA, digested with *Pst*I, *Hin*dIII, or *Eco*RI, hybridization signals yielded the expected size of fragments (data not shown). From these data as well as from the identical nucleotide sequence present in pCK3, it was concluded that the cloned *Hin*dIII fragments in pCK1 represent a contiguous *C. kluyveri* genomic DNA fragment and not a multiple ligation of *Hin*dIII fragments.

Nucleotide sequence analysis. For sequencing purposes and for localization of the 4-hydroxybutyrate dehydrogenase-encoding gene, nested deletion subclones were generated from pCK1 and pCK3 as described in Materials and Methods. The complete nucleotide sequence encompassing the insert of pCK1 (6,575 bp) was submitted to the EMBL data base (see Materials and Methods). DNA sequence analysis revealed four complete open reading frames (ORFs) (referred to as *orfY*, *cat1*, *sucD*, and *4hbD*) and two truncated ORFs at the 5′ and 3′ ends, respectively (*orfZ* and *sigL*), in the same transcriptional orientation (Fig. 2 and 3). An additional reading frame, encoding 36 amino acids, was located in the intergenic region between *orfZ* and *orfY* (bp 856 to 966), but it is likely that it has no coding function, since its G+C content (20.7%) was very low compared with that of the other ORFs (31.1 to 36.2%) and with that reported for genomic DNA (29.8%) from *C. kluyveri* (1). In addition, the nucleotide sequence downstream of the corresponding AUG start codon contained an inverted repeat with a calculated free energy of −66.5 kJ/mol (17), which could inhibit ribosome binding. Two other inverted repeats were found downstream of the *4hbD* gene. A free energy of −62.8 kJ/mol was calculated for the first repeat, which is 33 bp downstream of the *4hbD* stop codon (bp 6252 to 6285). The corresponding mRNA stem-loop would be followed by several U's as is typical for a rho-independent terminator (39). The second repeat had a calculated free energy of −44 kJ/mol and was located 133 bp further downstream. The AUG start codons of *cat1*, *sucD*, and *4hbD* were preceded by putative ribosome binding sites at a distance of 7, 6, and 7 bases, respectively, with reasonable homology to those described for *E. coli* (47) and for clostridia (60). In the case of the 3′-truncated ORF (*sigL*), the Shine-Dalgarno sequence was at a distance of 17 bases from the AUG codon. However, there was another possible start codon (GUG) in frame which would reduce the distance to the ribosome binding site (9 bases). For *orfY*, two possible sites for translation initiation were found. Upstream of the usual start codon, AUG (bp 1,050), with the putative ribosome binding site at a distance of 11 bp (AGGAG), there was a second possible start codon (UUG, bp 981), which also had a putative Shine-Dalgarno sequence at a distance of 8 bp (GGAGG).

Identification and expression of *4hbD*, *sucD*, and *cat1*. By analyzing the phenotype of the nested deletion subclones, the DNA region encoding the 4-hydroxybutyrate dehydrogenase gene was identified. Subclones having a deletion within the last complete ORF (*4hbD*) were no longer able to oxidize 4-hydroxybutyrate. That this gene (371 amino acids [aa], 41,755 Da) encodes the 4-hydroxybutyrate dehydrogenase was further confirmed by homologies to several class III alcohol dehydrogenases (see below).

The adjacent ORF upstream had an N-terminal amino acid sequence which was identical to that determined for the purified CoA-dependent succinate-semialdehyde dehydrogenase (M-[S]-N-E-V-S-I-K-E-L-I-E-K-A-K-V-A-Q-K-K-L-E-[A]-Y), except for one mismatch in position 16 where a valine was found in the protein but an alanine was deduced from the nucleotide sequence. Therefore, that ORF was designated *sucD*. For the *sucD* gene product (472 aa), a molecular mass of 50,915 Da was calculated, which was somewhat smaller than that determined for the purified enzyme by SDS-polyacrylamide gel electrophoresis (PAGE) (55 kDa [49]). CoA-dependent succinate-semialdehyde dehydrogenase activities in recombinant *E. coli* clones harboring the *sucD* gene (pCK1 to 4) were very low (10 to 20 mU/mg).

cat1, which was located directly upstream of *sucD*, encoded a 538-aa protein with a calculated molecular mass of 58,852 Da. A database search showed significant homologies (41.8 and 40.2% identity and 62.4 and 61.4% similarity) to an acetyl-CoA hydrolase (Ach1) from *Saccharomyces cerevisiae* (31) and to a gene product from *Neurospora crassa* (Acu-8) which is essential for growth on acetate (34), respectively. For bioenergetic reasons, we assume that in *C. kluyveri* this ORF encodes a CoA transferase rather than an acetyl-CoA-hydrolyzing enzyme. However, as with the CoA-dependent succinate-semialdehyde dehydrogenase, the specific activities of a succinyl-CoA:CoA transferase in recombinant *E. coli* clones were very low (5 to 16 mU/mg).

To induce transcription of *cat1*, *sucD*, and *4hbD* starting from the pBluescript *lac* promoter, recombinant clones harboring either a part (pCK169 [*'orfZ orfY cat1 sucD'*; bp 1 to 3772], pCK2 [*'cat1 sucD 4hbD sigL'*; bp 2117 to 6575]) or the complete gene region (pCK1, bp 1 to 6575) were grown in the presence of IPTG (1 mM), and cell extracts were prepared and analyzed for enzyme activity as described in Materials and Methods (Table 1). Cell extracts from *E. coli* harboring either pCK1 or pCK169 revealed significant succinyl-CoA:CoA transferase activity. Since the release of CoASH (monitored by the reduction of NbS_2) was dependent on all assay components (succinyl-CoA, oxaloacetate, citrate synthase, and acetate), a simple CoA-hydrolyzing activity like Ach1 from *S. cerevisiae* could be excluded. CoA-dependent succinate-semialdehyde dehydrogenase and 4-hydroxybutyrate dehydrogenase activities were present in *E. coli* (pCK1) and *E. coli* (pCK2) (Table 1). In the case of succinate-semialdehyde dehydrogenase, the assay was performed both in the presence and in the absence of CoA in order to differentiate between the CoA-independent *E. coli* enzyme, which occurred in all clones, and the recombinant CoA-dependent enzyme from *C. kluyveri* (Table 1). SDS-PAGE analysis during induction of the recombinant *E. coli* clones revealed an increase of protein bands corresponding to molecular masses of 66, 55, and 37 kDa (not shown). It is apparent that the 66-, 55-, and 37-kDa protein bands correspond to the *cat1*, *sucD*, and *4hbD* gene products. The 66-kDa protein occurred only in recombinant clones harboring the *cat1* gene (pCK1 and pCK169), whereas the 55- and 37-kDa proteins were present only in clones harboring the *sucD* and *4hbD* genes.

SucD and 4HbD sequence comparisons. A database search with the amino acid sequences of SucD (472 aa) and 4HbD (371 aa) revealed homologies to the *adhE* gene product from *E. coli* (891 aa) and to homologous proteins from two *Clostridium acetobutylicum* strains (DSM792, AdhE, 862 aa; ATCC 824, Aad, 873 aa) and an anaerobic protozoan (*Entamoeba histolytica*, Adh2, 870 aa), respectively (16, 20, 35, 57). The protein sequences from the *C. acetobutylicum* strains are almost identical and differ only with respect to the last 11 aa. Whereas the *C. kluyveri* SucD protein showed high similarity to the AdhE (Aad, Adh2) N-terminal region, the 4HbD protein corresponded to the C-terminal region (Table 2). The *E. coli adhE* gene product has both CoA-dependent aldehyde dehydrogenase activity and alcohol dehydrogenase activity (20). These functions are also proposed for the *adhE* and *aad* gene

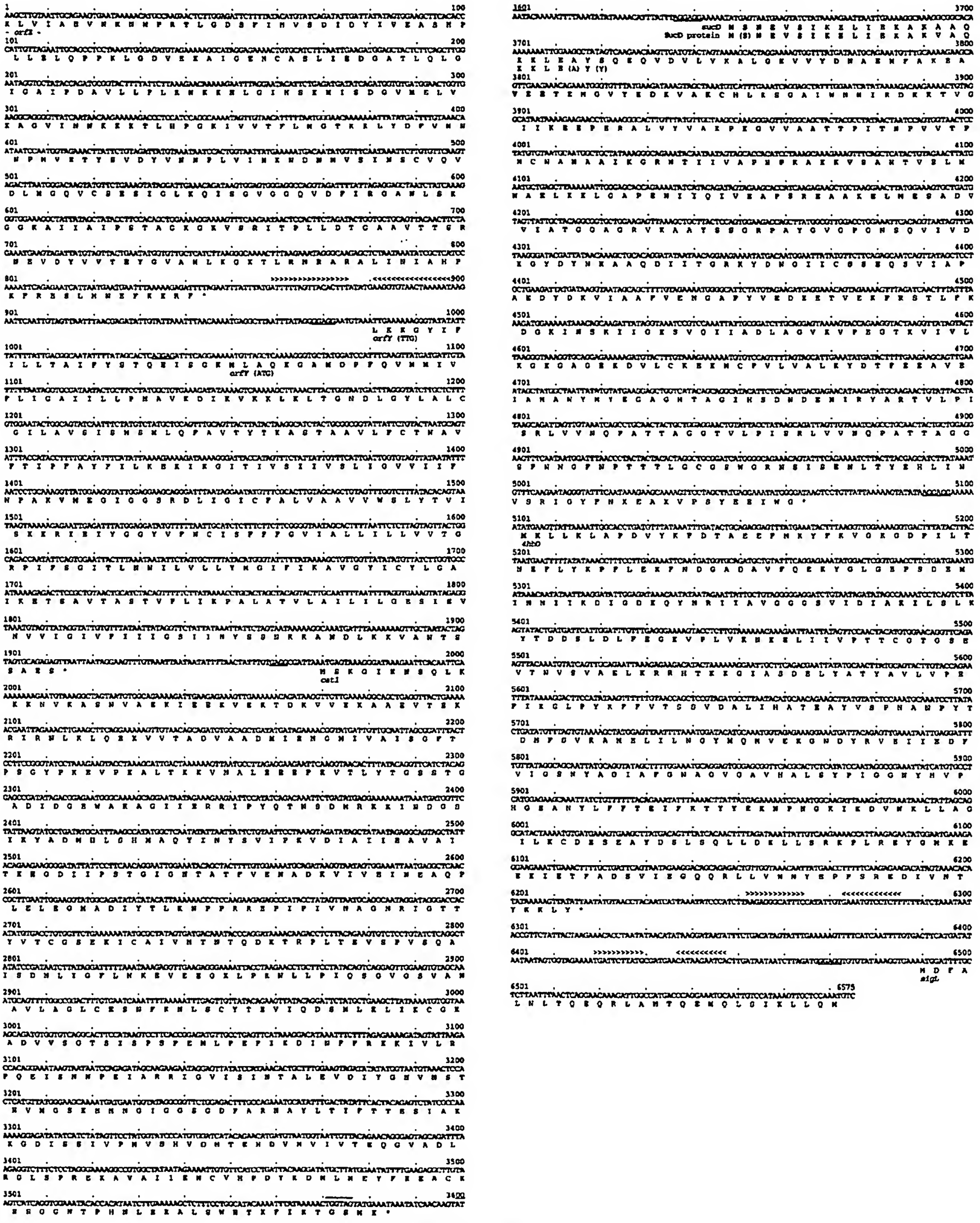

FIG. 3. Nucleotide sequence of the 6,575-bp insert of pCK1 representing the region of chromosomal DNA from *C. kluyveri* that contains *orfZ* (3′-terminal fragment), *orfY*, *catI*, *sucD*, *4hbD*, and *sigL* (5′-terminal fragment). The genes have been translated with the one-letter amino acid code with the symbols below the first nucleotide of the corresponding codon. Putative ribosome binding sites are underlined, and inverted repeats are indicated by open arrows representing the length and orientation of the stems. Bars above the sequence indicate the −35 and −10 regions of a putative promoter motif.

TABLE 1. Enzyme activities of recombinant *E. coli* clones harboring the *C. kluyveri cat1*, *sucD*, and *4hbD* genes after induction from the pBluescript *lac* promoter[a]

Strain and plasmid	Sp act (mU/mg)			
	Succinyl-CoA: CoA-transferase	Succinate-semialdehyde dehydrogenase		4-Hydroxybutyrate dehydrogenase
		−CoA[b]	+CoA[b]	
E. coli JM109 pBluescript	≤1	19 ± 1	≤1	≤2
E. coli JM109 pCK1 (*cat1 sucD 4hbD*)	184 ± 10	51 ± 3	80 ± 17	273 ± 66
E. coli JM109 pCK169 (*cat1*)	220 ± 13	23 ± 2	≤1	22 ± 5
E. coli JM109 pCK2 (*sucD 4hbD*)	≤1	49 ± 1	101 ± 15	475 ± 87

[a] Cells were grown in LB medium containing ampicillin (75 mg/liter), and IPTG (1 mM) was added at an optical density of 0.5. Crude extracts were prepared and analyzed for enzyme activity as described in Materials and Methods. Data represent average values and standard deviations from two independent experiments.

[b] −CoA and +CoA refer to the CoA-dependent and -independent succinate-semialdehyde dehydrogenase activities.

products from *C. acetobutylicum* DSM792 and ATCC 824, respectively (16, 35). Significant similarities were also found between the *C. kluyveri* 4HbD protein and other alcohol dehydrogenases (Table 2). These enzymes as well as the *E. coli* and *C. acetobutylicum adhE* (*aad*) gene products represent a new class (III) of alcohol dehydrogenases, which differ from both the long-chain zinc-containing (type I) and the short-chain zinc-lacking (type II) enzymes (2, 26, 40).

Analysis of *orfY*, *orfZ*, and *sigL*. As mentioned above, two possible sites for translation initiation of *orfY* were found. Hydropathy analysis of the deduced protein sequence revealed several highly hydrophobic regions. A translation start with AUG (as indicated by the arrow) would result in a protein of 288 amino acids (30,873 Da) with nine hydrophobic and probably membrane-spanning regions (Fig. 4). A translation start with UUG (311 aa, 33,507 Da) would produce a signal sequence with a basic N terminus (two lysine residues) followed by 12 hydrophobic amino acids. A ϕ(*lacZ'-'orfY*)*hyb30* fusion containing the first 30 aa of the *lacZ* α-peptide and 279 aa of *orfY* (starting with Pro-33 at bp 1073) was constructed from pCK1 with the *Bam*HI restriction endonuclease site (pCK7). Recombinant *E. coli* clones were able to grow in LB medium, but if expression of the fusion protein was induced by the addition of IPTG (1 mM), cell lysis occurred, probably because of a disintegration of the cell membrane (not shown).

Database searches revealed similarities (about 50%) to other putative membrane-spanning proteins from *Desulfurolobus ambivalens* (29), *Haemophilus influenzae* (52), and *E. coli* (DDBJ:D13267). However, the function of these proteins is yet to be determined, and most of the similarities between them might be due to their common membrane-spanning function.

Whereas the amino acid sequence deduced from the 5'-truncated ORF *orfZ* showed no significant homology to protein sequences in the databases, the N-terminal amino acid sequence deduced from the last reading frame (*sigL*), which was located at the 3' end of the pCK1 insert, revealed significant similarities to the family of σ^{54}-related proteins (not shown). Since the N terminus of these proteins is highly conserved, it seems likely that *sigL*, which was designated in analogy to the corresponding gene from another gram-positive organism (*Bacillus subtilis*), encodes a σ^{54}-homologous sigma factor.

Northern blot analysis. To analyze the expression of the cloned genes in *C. kluyveri*, RNA was isolated from cells grown on ethanol plus succinate and ethanol plus acetate and prepared for Northern blot hybridization. With radiolabeled probes complementary to the *4hbD* (bp 5668 to 6025) and the *cat1* (bp 2117 to 2981) genes, significant hybridization predominantly occurred with RNA isolated from cells grown on ethanol plus succinate (Fig. 5). A strong signal corresponding to approximately 2,700 nucleotides in length was obtained with the *4hbD*-complementary probe. In addition, two weak signals of about 5,500 and 9,500 nucleotides and a small transcript of about 1,800 nucleotides could be detected after prolonged exposure (Fig. 5). The *cat1*-complementary probe gave similar results: only low hybridization occurred with RNA from ethanol-acetate-grown cells, whereas with RNA from cells grown on ethanol plus succinate, signals corresponding to a length of

TABLE 2. Comparison of the *C. kluyveri* gene products SucD (succinate-semialdehyde dehydrogenase) and 4HbD (4-hydroxybutyrate dehydrogenase) with aldehyde and alcohol dehydrogenases from other organisms

Organism and gene product (reference[s])	Sequence positions[a]	*C. kluyveri* SucD[b]		*C. kluyveri* 4HbD[b]	
		Identity (%)	Similarity (%)	Identity (%)	Similarity (%)
C. acetobutylicum AdhE-Aad (16, 35)	2–458	41.8	58.8		
	449–860			26.1	51.9
E. coli AdhE (20)	3–474	37.2	57.7		
	451–862			26.2	51.4
Entamoeba histolytica AdhE2 (57)	7–491	38.0	60.8		
	462–862			26.1	51.9
C. acetobutylicum Adh1 (61)	1–381			28.2	55.2
Saccharomyces cerevisiae Adh4 (54)	4–380			26.4	51.4
Zymomonas mobilis ATCC 10988 AdhB (59)	1–383			27.0	50.1
Zymomonas mobilis ZM4 AdhB (11)	1–383			26.6	49.7
Citrobacter freundii (12)	25–393			26.2	50.0
Bacillus methanolicus Mdh (14)	3–381			25.9	49.3
E. coli FucO (10)	5–383			23.6	51.9
C. acetobutylicum BdhA (51)	6–389			22.9	47.9
C. acetobutylicum BdhB (51)	18–388			21.2	46.9

[a] Numbers refer to the first and last amino acids of the resulting sequence comparison.

[b] Sequence comparison was performed with the Bestfit algorithm (Genetics Computer Group package [13]) with a gap weight of 3.0 and a length weight of 0.1.

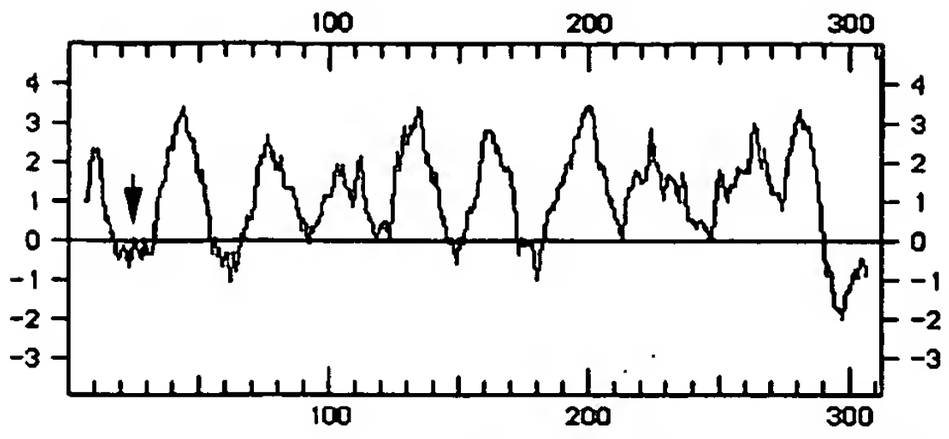

FIG. 4. Hydropathy plot of the *orfY* gene product according to the work of Kyte and Doolittle (30). The amino acid sequence was deduced from a translation start with UUG (bp 981 [Fig. 3]), and the arrow indicates the first amino acid from a translation start with AUG (bp 1050 [Fig. 3]). Positive values represent high hydrophobicity and negative values indicate low hydrophobicity, averaged over a window of 7 aa (30).

about 2,700, 5,500, and 9,500 nucleotides and some small transcripts were obtained (Fig. 5).

DISCUSSION

We screened recombinant *E. coli* clones for an oxidative utilization of 4-hydroxybutyrate and could isolate a DNA region from *C. kluyveri* encoding the 4-hydroxybutyrate dehydrogenase, the CoA-dependent succinate-semialdehyde dehydrogenase, and a succinyl-CoA:CoA transferase. Since *E. coli* has two CoA-independent succinate semialdehyde dehydrogenases which directly catalyze the oxidation of succinate semialdehyde to succinate (15), the use of 4-hydroxybutyrate as a carbon and energy source by *E. coli* depends only on the presence and expression of the clostridial 4-hydroxybutyrate dehydrogenase-encoding gene (*4hbD*) as shown by the nested deletion subclone analysis.

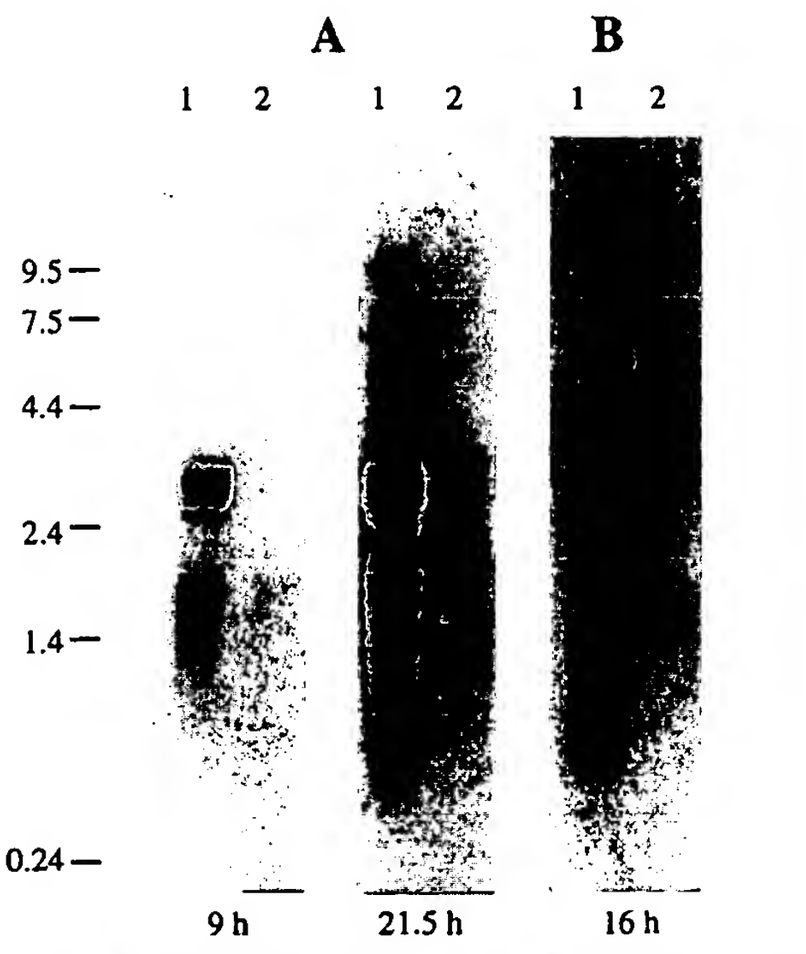

FIG. 5. Northern blot hybridization with radiolabeled fragments which are complementary to the *C. kluyveri 4hbD* (A) and *cat1* (B) genes. Lanes 1, 8 μg of total RNA from *C. kluyveri* cells, grown on ethanol plus succinate; lanes 2, 8 μg of total RNA from *C. kluyveri* cells, grown on ethanol plus acetate. The cells were harvested in the logarithmic growth state. The sizes of selected marker bands (RNA ladder) are indicated in kilobases at the left side of the figure. (A) Results obtained after hybridization with a *4hbD*-complementary probe and autoradiography for 9 and 21.5 h. (B) The blot was hybridized against a *cat1*-complementary probe and subjected to autoradiography for 16 h, as indicated.

Thus far, the only known protein sequences of CoA-acylating bacterial aldehyde dehydrogenases were those from the multifunctional AdhE enzymes from *E. coli* and the homologous gene from *C. acetobutylicum*, which have both CoA-dependent aldehyde dehydrogenase and an alcohol dehydrogenase activity (16, 20, 35). From the tandem arrangement of the *C. kluyveri sucD* and *4hbD* genes and the similarities to the N- and C-terminal regions of the *E. coli* and *C. acetobutylicum adhE* (*aad*) gene products, it is evident that the AdhE proteins are composed of two catalytic domains: the N terminus exhibits a CoA-acylating aldehyde dehydrogenase activity whereas the C-terminal part is an alcohol dehydrogenase. This has also been suggested by Goodlove et al. (20), Fischer et al. (16), and Nair et al. (35). However, it should be noted that the *E. coli* AdhE protein has a third function, the deactivation of the pyruvate:formate-lyase (28). Primary protein sequence comparison of the *C. kluyveri* SucD-4HbD proteins and the *adhE* (*aad*) gene products of *C. acetobutylicum*, *E. coli*, and the protozoan *E. histolytica* reveals some interesting features (Fig. 6). The N-terminal region which corresponds to the *C. kluyveri* SucD protein (positions 1 to 472) allows now a first comparison of CoA-acylating aldehyde dehydrogenases (Fig. 6A). The sequence is highly conserved: from a consensus length of 490, 138 aa are strictly conserved in the bacterial enzymes (99 aa including the protozoan AdhE protein sequence) with 321 aa being present in two of the three microbial species. The SucD homologous sequence includes one highly conserved motif for nucleotide binding (G-x-G-x-x-G; site 4) (53) and another region with G-rich motifs (G-G-x-G; G-x-G-x-G; site 2), one of which also might contribute to ADP binding. Since CoA, like $NAD(P)^+$, contains the ADP moiety, two nucleotide binding sites are reasonable. Another highly conserved region (site 3) contains a glycine and a cysteine residue and shows limited similarity to the active center proposed for CoA-independent aldehyde dehydrogenases (2, 23). It is suggested for these enzymes as well as for the CoA-acylating aldehyde dehydrogenases that an enzyme-bound thiol group (cysteine) is required at the catalytic site for the formation of a hemithioacetal intermediate (25, 48). Database searches performed with the AdhE (Aad) proteins from *C. acetobutylicum* revealed significant but low homology to aldehyde dehydrogenases (CoA independent) from different sources (<25% identity). In addition to the active center described above, a conserved dodecapeptide of CoA-independent aldehyde dehydrogenases (G-V-TC-TGV-GQ-I-LIS-P-W-N-FY-P [24]) would correspond to positions 111 to 122 of the CoA-dependent enzymes (site 1 [Fig. 6A]). The proposed nucleotide binding site of CoA-independent aldehyde dehydrogenases would match positions 174 to 179 and is thus not conserved. In addition, a typical decapeptide (V-TC-L-E-L-G-G-K-AS-P) of these enzymes is missing. Since there is no similarity of CoA-dependent and CoA-independent aldehyde dehydrogenases in the C-terminal region, the highly conserved nucleotide binding site of CoA-dependent enzymes (site 4) might be involved in CoA binding.

The *C. kluyveri* 4HbD protein as well as the AdhE proteins belongs to the class III alcohol dehydrogenases (2, 26, 40). Sequence comparison revealed a more or less common motif, which comprises several histidine residues (Fig. 6B, sites 1 and 2) and might be involved in divalent cation (iron) binding (2, 16, 61). The *C. kluyveri* 4HbD protein contains three of the four histidine residues proposed to be involved in iron binding. However, a glutamate and an aspartate residue (site 1) are

A

```
CkSucD MSNEVSIKELIEKAKAAQKKLE--------AYSQEQVDVLVKALGKVVYDNAEMFAKEAVEETEMGVYEDKVAKCHLKSGAIWNHIKDKKTVGIIKEDPE  92
CaAdhE MKVTTVKELDEKLKVIKEAQKKF------SCYSQEMVDEIFRNAAMAAIDARIELAKAAVLETGMGLVEDKVIKNHFAGEYIYNKYKDEKTCGIIERNEP  94
CaAad  MKVTTVKELDEKLKVIKEAQKKF------SCYSQEMVDEIFRNAAMAAIDARIELAKAAVLETGMGLVEDKVIKNHFAGEYIYNKYKDEKTCGIIERNEP  94
EcAdhE MAVTNVAELNALVERVKKAQREY------ASFTQEQVDKIFRAAALAAADARIPLAKMAVAESGMGIVEDKVIKNHFASEYIYNAYKDEKTCGVLSEDDT  94
EhAdhE MSTQQTMTVDEHINQLVRKAQVALKEYLKPEYTQEKIDYIVKKASVAALDQHCALAAAAVEETGRGIFEDKATKNIFACEHVTHEMRHAKTVGIINVDPL 100
                .       .       .       .       .       .       .       .       . 100

                 GV=====P===P                                                  G
CkSucD RALVYVAKPKGVVAATTPITNPVVTPMCNAMAAIKGRNTIIVAPHPKAKKVSAHTVELMNAELKKLGAPENIIQIVEAPSREAAKELMESADV--VIATG 190
CaAdhE YGITKIAEPIGVVAAIIPVTNPTSTTIFKSLISLKTRNGIFFSPHPRAKKSTILAAKTILDAAVKSGAPENIIGWIDEPSIELTQYLMQKADITL--ATG 192
CaAad  YGITKIAEPIGVVAAIIPVTNPTSTTIFKSLISLKTRNGIFFSPHPRAKKSTILAAKTILDAAVKSGAPENIIGWIDEPSIELTQYLMQKADITL--ATG 192
EcAdhE FGTITIAEPIGIICGIVPTTNPTSTAIFKSLISLKTRNAIIFSPHPRAKDATNKAADIVLQAAIAAGAPKDLIGWIDQPSVELSNALMHHPDINLILATG 194
EhAdhE YGITEIAEPVGVVCGVTPVTNPTSTAIFKSLISIKTRNPIVFSFHPSALKCSIMAAKIVRDAAIAAGAPENCIQWIEFGGIEASNKLMNHPGVATILATG 200
                 L=======1=======J       .       .       .       .       .       L= 200

       G=G=========G=====G=G=G=]                 [=G=C=========]
CkSucD GAGRVKAAYSSGRPAYGVGPGNSQVIVDKGYDYNKAAQDIITGRKYDNGIICSSEQSVIAPAEDYDKVIAAFVENGAFYVEDEETVEKFRSTL------- 283
CaAdhE GPSLVKSAYSSGKPAIGVGPGNTPVIIDESAHIKMAVSSIILSKTYDNGVICASEQSVIVLKSIYNKVKDEFQERGAYIIKKNELDKVREVIF------- 285
CaAad  GPSLVKSAYSSGKPAIGVGPGNTPVIIDESAHIKMAVSSIILSKTYDNGVICASEQSVIVLKSIYNKVKDEFQERGAYIIKKNELDKVREVIF------- 285
EcAdhE GPGMVKAAYSSGKPAIGVGAGNTPVVIDETADIKRAVASVLMSKTFDNGVICASEQSVVVVDSVYDAVRERFATHGGYLLQGKELKAVQDVIL------- 287
EhAdhE GNAMVKAAYSSGKPALGVGAGNVPTYIEKTCNIKQAANDVVMSKSFDNGMICASEQAAIIDKEIYDQVVEEMKTLGAYFINEEEKAKLEKFMFGVNAYSA 300
       ==========2==========J       .       .       L=========3=========J       .       .       . 300

CkSucD -FKDGKINSKIIGKSVQIIADLAGVKVPEGTKVIVLKGKGAGEKDVLCKEKMCPVLVALKYDTFEEAVEIAMANYMYEGAGHTAGIHSDNDENIRYARTV 382
CaAdhE --KDGSVNPKIVGQSAYTIAAMAGIKVPKTTRILIGEVTSLGEEEPFAHEKLSPVLAMYEADNFDDALKKAVTLINLGGLGHTSGIYADEIKA------- 376
CaAad  --KDGSVNPKIVGQSAYTIAAMAGIKVPKTTRILIGEVTSLGEEEPFAHEKLSPVLAMYEADNFDDALKKAVTLINLGGLGHTSGIYADEIKA------- 376
EcAdhE --KNGALNAAIVGQPAYKIAELAGFSVPENTKILIGEVTVVDESEPFAHEKLSPTLAMYRAKDFEDAVEKAEKLVAMGGIGHTSCLYTDQDNQ------- 378
EhAdhE DVNNARLNPKCPGMSPQWFAEQVGIKVPEDCNIICAVCKEVGPNEPLTREKLSPVLAILKAENTQDGIDKAEAMVEFNGRGHSAAIHSND---------- 390
                .       .       .       .       .       .       .       .       . 400

                                                 [=G=G=G=]
CkSucD LPISRLVVNQPATTAGGTVLPISRLVVNQPATTAGGSFNNGFNPTTTLGCGSWGRNSISENLTYEHLINVSRIGYFNKEAKVPSYEEIWG 472
CaAdhE -------RDKIDRFSSAMKTVRTFVNIPTSQGASGDLYNFRIPPSFTLGCGFWGGNSVSENVGPKHLLNIKTVAERRENMLWFRVPHKVY 459
CaAad  -------RDKIDRFSSAMKTVRTFVNIPTSQGASGDLYNFRIPPSFTLGCGFWGGNSVSENVGPKHLLNIKTVAERRENMLWFRVPHKVY 459
EcAdhE -------PARVSYFGQKMKTARILINTPASQGGIGDLYNFKLAPSLTLGCGSWGGNSISENVGPKHLINKKTVAKRAENMLWHKLPKSIY 461
EhAdhE -------KAVVEKYALTMKACRILHNTPSSQGGIGSIYNY-IWPSFTLGCGSYGGNSVSANVTYHNLLNIKRLADRRNNLQWFRVPPKIF 472
                .       .       .       .       L=======4=======J       .       .       . 490
```

B

```
Ck4HbD MKLLKLAPDVYKFDTAEEFMKYFKVGKGDFILTNEFLYKPFLEKFNDGADAVFQEKYGLGEPSDEMINNIIKDIGDKQYNRIIAVGGGSVIDIAKILSLK 100
CaAdhE FKFGCLQFALKDLKDLKKKRAFIVTDSDPYNLNYVDSIIKILEHL--DIDFKVFNKVG-READLKTIKKATEEMSSFMPDTIIALGGTPEMSSAKLMWVL 556
CaAad  FKFGCLQFALKDLKDLKKKRAFIVTDSDPYNLNYVDSIIKILEHL--DIDFKVFNKVG-READLKTIKKATEEMSSFMPDTIIALGGTPEMSSAKLMWVL 556
EcAdhE FRRGSLPIALDEVITDGHKRALIVTDRFLFNNGYADQITSVLKAA--GVETEVFFEVE-ADPTLSIVRKGAELANSFKPDVIIALGGGSPMDAAKIMWVM 558
EhAdhE FEPHSIRYLAELKE---LSKIFIVSDRMMYKLGYVDRVMDVLKRRSNEVEIEIFIDVE-PDPSIQTVQKGLAVMNTFGPDNIIAIGGGSAMDAAKIMWLL 568
                .       .       .       .       .       .       .       .       . 100

                                                                    [=D=====H=
Ck4HbD YTDDSLDLFEGKVPLV-------------KNKELIIVPTTCGTGSEVTNVSVAELKRRHTKKGIASDELYATYAVLVPEFIKGLPYKFFVTSSVDALIHA 187
CaAdhE YEHPEVKFEDLAIKFMDIRKRIYTFPKLGKKAMLVAITTSAGSGSEVTPFALVTDNNTGNKYMLADYEMTPNMAIVDAELMMKMPKGLTAYSGIDALVNS 656
CaAad  YEHPEVKFEDLAIKFMDIRKRIYTFPKLGKKAMLVAITTSAGSGSEVTPFALVTDNNTGNKYMLADYEMTPNMAIVDAELMMKMPKGLTAYSGIDALVNS 656
EcAdhE YEHPETHFEELALRFMDIRKRIYKFPKMGVKAKMIAVTTTSGTGSEVTPFAVVTDDATGQKYPLADYALTPDMAIVDANLVMDMPKSLCAFGGLDAVTHA 658
EhAdhE YEHPEADFFAMKQKFIDLRKRAFKFPTMGKKARLICIPTTSGTGSEVTPFAVISDHETGKKYPLADYSLTPSVAIVDPMFTMSLPKRAIADTGLDVLVHA 668
                .       .       .       .       .       .       .       .       L=====1= 200

       =E=]                                                [=G=H=AH=G=======PHG=]
Ck4HbD TEAYVSPNANPYTDMFSVKAMELILNGYMQMVEKGNDYRVEIIEDFVIGSNYAGIAFGNAGVGAVHALSYPIGGNYHVPEGEANYLFFTEIFKTYYEKNP 277
CaAdhE IEAYTSVYASEYTNGLALEAIRLIFK-YLPEAYKNGRTNEKAREKMAHASTMAGMASANAFLGLCHSMAIKLSSEHNIPSGIANALLIEEVIKFNAVDNP 755
CaAad  IEAYTSVYASEYTNGLALEAIRLIFK-YLPEAYKNGRTNEKAREKMAHASTMAGMASANAFLGLCHSMAIKLSSEHNIPSGIANALLIEEVIKFNAVDNP 755
EcAdhE MEAYVSVLASEFSDGQALQALKLLKE-YLPASYHEGSKNPVARERVHSAATIAGIAFANAFLGVCHSMAHKLGSQFHIPHGLANALLICNVIRYNANDNP 757
EhAdhE TEAYVSVMANEYTDGLAREAVKLVFE-NLLKSYNG---DLEAREKMHNAATIAGMAFASAFLGMDHSMAHKVGAAFHLPHGRCVAVLLPHVIRYNG-QKP 763
       L=J       .       .       .       .       .       L=========2=========J       .       . 300

Ck4HbD ------------NGKIKDVNKLLAGILKCDESEAYDSLSQLLDKLLSRKPLREYGMKEEEIETFADSVIEGQQRLLVNNYEP----------------- 347
CaAdhE VKQAPCPQYKYPNTIFRYARIADYIKLG--GNTDEEKVDLLINKIHELKKALNIPTSIKDAGVLEENFYSSLDRISELALDDQCTGANPRFPLTSEIKEM 846
CaAad  VKQAPCPQYKYPNTIFRYARIADYIKLG--GNTDEEKVDLLINKIHELKKALNIPTSIKDAGVLEENFYSSLDRISELALDDQCTGANPRFPLTSEIKEM 853
EcAdhE TKQTAFSQYDRPQARRRYAEIADHLGLSAPGDRTAAKIEKLLAWLETLKAELGIPKSIREAGVQEADFLANVDKLSEDAFDDQCTGANPRYPLISELKQI 857
EhAdhE RKLAMWPKYNFYKADQRYMELAQMVGLKC--NTPAEGVEAFAKACEELMKATETITGFKKANIDEAAWMSKVPEMALLAFEDQCSPANPRVPMVKDMEKI 858
                .       .       .       .       .       .       .       .       . 400

Ck4HbD --------------------FSREDIVNTYKKLY 371
CaAdhE YIN----------CF---------------KKQP 862
CaAad  YINFVLKNNLKPSYP---------------NYFL 872
EcAdhE LLDTYYGRDYVEGETAAKKEAAPAKAEKKAKKSA 891
EhAdhE LKAAYYP-------------------------IA 870
                .       .       .     924
```

FIG. 6. Amino acid alignment of the *sucD* and *4hbD* gene products from *C. kluyveri* (CkSucD [A] and Ck4HbD [B], respectively) with the *adhE* gene products from *C. acetobutylicum* (CaAdhE [16] and CaAad [35]), *E. coli* (EcAdhE [20]), and *E. histolytica* (EhAdhE [57]). The *C. kluyveri* SucD and 4HbD proteins are depicted in the first line and cover the positions 1 to 472 and 1 to 371, as indicated. (A) The first site indicates a conserved dodecapeptide for aldehyde dehydrogenases (24). Glycine-rich motifs (sites 2 and 4) indicate putative nucleotide binding sites of the aldehyde dehydrogenase domain. Site 3 depicts the proposed active center of the aldehyde dehydrogenase domain with a conserved cysteine residue. (B) Sites 1 and 2 represent the proposed iron binding motif for the class III alcohol dehydrogenases (2, 61) comprising several histidine residues.

```
CkCat1 MSKGIKNSQLKKKNVKASNVAEKIEEKVEKTDKVVEKAAEVTEKRIRNLKLQEKVVTADVAADMIENGMIVAISGFTPSGYPKEVPKALTKKVNALEEE-  99
NcAcu8 MASPI---------------------------------ASAALRARQGPSMLKKLCNPEDMLQHFPNGAYIGWSGFTASATPRR-SYYLADHVEKNASRP  66
ScAch1 MT--I---------------------------------SNLLKQRVRYAPYLKKVKEAHELIPLFKNGQYLGWSGFTGVGTPKAVPEALIDHVEKNNLQG  65
                .       .       .       .       .       .       .       .       . 100

CkCat1 ---FKVTLYTGSSTGADIDGEWAKAGIIERRIPYQTNSDMRKKINDGSIKYADMHLSHMAQYINYSVIPK-------VDIAIIEAVAITEEGDIIPSTGI 189
NcAcu8 AQVQPLRRCLRRRRDRERWAALDMIAHDRRRAPHQVGKNIAKGINEGRINFFDKHLSMFPVDLVYGYYTKDR-QNKNLDVVCVEATEIKEDGSIVLGASV 165
ScAch1 KLRFNLFVGASAGPEENRWAEHDMII---KRAPHQVGKPIAKAINQGRIEFFDKHLSMFPQDLTYGFYTRERKDNKILDYTIIEATAIKEDGSIVPGPSV 162
                .       .       .       .       .       .       .       .       . 200

CkCat1 GNTATFVENADKVIVEINEAQPLELEGMADIYTLKNPPRREPIPIVNAGNRIGTTYVTCGSEKICAIVMTNTQDKTRPLTEVSPVSQAISDNLIGFLNKE 289
NcAcu8 GATPELIQMADKVIIEVNTAIP-SFDGLHDITFSDLPPNRKPYLIQQCRDRIGTTSVPVDPEKVVGIIECTTPDQTLPNSPADETATAIAGHLIEFFEHE 264
ScAch1 GGSPEFITVSDKVIIEVNTATP-SFEGIHDIDMPVNPPFRKPYPYLKVDDKCGVDSIPVDPEKVVAIVESTMRDQVPPNTPSDDMSRAIAGHLVEFFRNE 261
                .       .       .       .       .       .       .       .       . 300

CkCat1 VEEGKLPKNLLPIQSGVGSVANAVLAGLCESNFKNLSCYTEVIQDSMLKLIKCGKADVVSGTSISPSPEMLPEFIKDINFFREKIVLRPQEISNNPEIAR 389
NcAcu8 VAHGRLPKNLLPLQSGIGNIANAVIGGLETSNFKNLNVWTEVIQDTFLDLFDSGKLDFATATSIRFSPTGFERFYKNWDNYYDKLLLRSQSVSNAPEIIR 364
ScAch1 VKHGRLPENLLPLQSGIGNIANAVIEGLAGAQFKHLTVWTEVLQDSLLDLFENGSLDYSTATSVRLTEKGFDRAFANWENFKHRLCLRSQVVSNNPEMIR 361
                .       .       .       .       .       .       .       .       . 400

CkCat1 RIGVISINTALEVDIYGNVNSTHVMGSKMMNGIGGSGDFARNAYLTIFTTESI----AKKGDISSIVPMVSHVDHTEHDVMVIVTEQGVADLRGLSPREK 485
NcAcu8 RLGVIGMNTPVEVDIYAHANSTNVMGSRMLNGLGGSADFLRNSKYSIMHTPSTRPSKTDAHGVSCIVPMCTHVDQTEHDLDVIVTENGLADVRGLSPRER 464
ScAch1 RFPVIAMNTPVEVDIYAHANSTNVNGSRMLNGLGGSADFLRNAKLSIMHAPSARPTKVDPTGISTIVPMASHVDQTEHDLDILVTDQGLADLRGLSPKER 461
                .       .       .       .       .       .       .       .       . 500

CkCat1 AVAIIENCVHPDYKDMLMEYFEEAC----KSSGGNTPHNLEKALSWHTKFIKTGSMK-------- 538
NcAcu8 ARVIIDKCAHDVYKPILKAYFEKAEFECLRKGMGHEPHLLFNSFDMHKALVEEGSMAKVKF---- 525
ScAch1 AREIINKCAHPDYQALLTDYLDRAEHYAKKHNCLHEPHMLKNAFKFHTNLAEKGTMKVDSWEPVD 526
                .       .       .       .       .       .       565
```

FIG. 7. Amino acid alignment of the *C. kluyveri* succinyl-CoA:CoA transferase (CkCat1) with a protein necessary for acetate utilization in *N. crassa* (NcAcu8) and with an acetyl-CoA hydrolase from *S. cerevisiae* (ScAch1). Blocks indicate putative sites for nucleotide (CoA) binding.

strictly conserved among all class III alcohol dehydrogenases and might also be involved in metal binding. It is evident that the clostridial 4HbD enzyme requires divalent cations for activity; our assay includes Mg^{2+} (1 mM) as described elsewhere (22, 49). However, the ion specificity of the enzyme has not yet been investigated.

The *E. coli* AdhE is a homodimeric protein composed of two 96-kDa subunits (9). Interestingly, an AdhE-like aldehyde-alcohol dehydrogenase enzyme complex has also been purified and characterized from *C. kluyveri*, which is composed of two 55- and two 42-kDa subunits (33). From the data presented here, a gene arrangement similar to *sucD-4hbD* might be expected. Cloning and sequencing of the corresponding genes could be an important step for the elucidation of evolutionary relationships within this class of enzymes. A dimeric aldehyde dehydrogenase domain also corresponds to the composition of several CoA-acylating enzymes: the succinate-semialdehyde dehydrogenase from *C. kluyveri* (2 × 55 kDa [49]), the butyraldehyde dehydrogenase from *C. acetobutylicum* (2 × 55 kDa [36]), and the aldehyde dehydrogenase from *Clostridium beijerinckii* (2 × 56 kDa [56]).

During growth on ethanol and succinate, a succinyl-CoA:CoA transferase reaction is required for the initial activation of the substrate (49). Heterologous expression of the clostridial *cat1* gene in *E. coli* from the pBluescript *lac* promoter identified its gene product as a succinyl-CoA:CoA transferase. Northern blot analysis revealed that *cat1* is efficiently transcribed only during growth on ethanol plus succinate. We could not detect sequence similarities between Cat1 and the heterodimeric CoA transferases from *Pseudomonas putida*, *Acinetobacter calcoaceticus*, and *C. acetobutylicum* or the monomeric enzyme from pig heart, which is presumably cleaved after translation (18, 32, 37). It is somewhat surprising that the *C. kluyveri* succinyl-CoA:CoA transferase (Cat1) and the Acu-8 protein sequence from *N. crassa* indicate high homologies to an acetyl-CoA hydrolase from *S. cerevisiae* (Fig. 7). However, the conserved CoA (ADP) binding site from the heterodimeric CoA transferases (-G-x-G-x-x-G- [37]) is also abundant (-G-x-G-x-x-G/A-) in the amino acid sequences of Ach1 (*S. cerevisiae*), Acu-8 (*N. crassa*), and Cat1 (*C. kluyveri* [Fig. 7]). As in *C. kluyveri*, growth of *N. crassa* on acetate should not involve an acetyl-CoA hydrolase, but a succinyl-CoA:CoA transferase, which catalyzes the activation of acetate, might be required.

Northern blot analysis revealed that the cloned gene region is efficiently transcribed during growth on ethanol plus succinate (Fig. 5). The signal of 2,700 nucleotides obtained with the *4hbD*-complementary probe most likely represents a common transcription of *sucD* and *4hbD*, which together are 2,590 bp in length. The presence of a rho-independent terminator and an additional stem-loop downstream of the *4hbD* gene further supports this hypothesis. The signals of 5,500 and 9,500 nucleotides might represent a common transcription of *orfY*, *cat1*, *sucD*, and *4hbD* (which together cover 5,042 bp) and point to an additional transcription start point upstream of the cloned gene region. Because of the high A+T content of the clostridial DNA, especially in the intergenic regions, identification of putative promoter structures by sequence comparison alone is often ambiguous. However, a more or less significant motif (TGGTAG-21 bp-TATAAT) is located 44 bp upstream of *sucD*. Clearly, more experiments (including primer extension) are needed for a detailed transcription analysis.

The data on heterologous expression of *cat1*, *sucD*, and *4hbD* are in good agreement with the results from the work of Wolff et al. (55), who recently reported on several protein bands in cell extracts from *C. kluyveri*, which occurred only during growth on ethanol plus succinate (100, 66, 58, 55, 37, and 30 kDa). The 30-kDa protein band would correspond to *orfY*, but it could not be identified during expression of the cloned genes in *E. coli*, probably because of poor translation initiation.

ACKNOWLEDGMENTS

We thank B. Schmidt (Göttingen, Germany) for determination of the N-terminal amino acid sequence of the purified succinate-semialdehyde dehydrogenase and P. Dürre and R.-J. Fischer (Göttingen) for stimulating discussions.

This work was supported by a grant of the Akademie der Wissenschaften, Göttingen, Germany.

REFERENCES

1. **Bader, J., H. Günther, H. Simon, S. Pohl, and W. Mannheim.** 1980. Utilization of (E)-2-butenoate (crotonate) by *Clostridium kluyveri* and some other *Clostridium* species. Arch. Microbiol. 125:159–165.
2. **Bairoch, A.** 1991. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. 19(Suppl.):2241–2245.
3. **Barker, H. A.** 1942. *Clostridium kluyveri*, an organism concerned in the formation of caproic acid from ethyl alcohol. J. Biol. Chem. 43:347–363.
4. **Bartsch, R. G., and H. A. Barker.** 1961. A vinylacetyl isomerase from *Clostridium kluyveri*. Arch. Biochem. Biophys. 92:122–132.
5. **Bergmeyer, H. U., G. Holz, H. Klotzsch, and G. Lang.** 1963. Phosphotransacetylase aus *Clostridium kluyveri*. Züchtung des Bakteriums, Isolierung, Kristallisation und Eigenschaften des Enzyms. Biochem. Z. 338:114–121.
6. **Bochner, B. R., and M. A. Savageau.** 1977. Generalized indicator plate for genetic, metabolic, and taxonomic studies with microorganisms. Appl. Environ. Microbiol. 33:434–444.
7. **Bornstein, B. T., and H. A. Barker.** 1948. The energy metabolism of *Clostridium kluyveri* and the synthesis of fatty acids. J. Biol. Chem. 172:659–669.
8. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248–254.
9. **Clark, D. P.** 1989. The fermentation pathways of *Escherichia coli*. FEMS Microbiol. Rev. 63:223–234.
10. **Conway, T., and L. O. Ingram.** 1989. Similarity of *Escherichia coli* propanediol oxidoreductase (*fucO* product) and an unusual alcohol dehydrogenase from *Zymomonas mobilis* and *Saccharomyces cerevisiae*. J. Bacteriol. 171:3754–3759.
11. **Conway, T., G. W. Sewell, Y. A. Osman, and L. O. Ingram.** 1987. Cloning and sequencing of the alcohol dehydrogenase II gene from *Zymomonas mobilis*. J. Bacteriol. 169:2591–2597.
12. **Daniel, R., and G. Gottschalk.** Purification of 1,3-propanediol dehydrogenase from *Citrobacter freundii*: sequencing and overexpression of the corresponding gene in *Escherichia coli*. J. Bacteriol. 177:2151–2156.
13. **Devreux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387–395.
14. **De Vries, G. E., N. Arfman, P. Terpstra, and L. Dijkhuizen.** 1992. Cloning, expression, and sequence analysis of the *Bacillus methanolicus* C1 methanol dehydrogenase gene. J. Bacteriol. 174:5346–5353.
15. **Donnelly, M. I., and R. A. Cooper.** 1981. Two succinic semialdehyde dehydrogenases are induced when *Escherichia coli* K-12 is grown on γ-aminobutyrate. J. Bacteriol. 145:1425–1427.
16. **Fischer, R. J., J. Helms, and P. Dürre.** 1993. Cloning, sequencing and molecular analysis of the *sol* operon of *Clostridium acetobutylicum*, a chromosomal locus involved in solventogenesis. J. Bacteriol. 175:6959–6969.
17. **Freier, S. M., R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner.** 1986. Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. USA 83:9373–9377.
18. **Gerischer, U., and P. Dürre.** 1990. Cloning, sequencing, and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum*. J. Bacteriol. 172:6907–6918.
19. **Gerischer, U., and P. Dürre.** 1992. mRNA analysis of the *adc* gene region of *Clostridium acetobutylicum* during the shift to solventogenesis. J. Bacteriol. 174:426–433.
20. **Goodlove, P. E., P. R. Cunningham, J. Parker, and D. P. Clark.** 1989. Cloning and sequence analysis of the fermentative alcohol-dehydrogenase-encoding gene of *Escherichia coli*. Gene 85:209–214.
21. **Gottschalk, G.** 1986. Bacterial metabolism, 2nd ed. Springer Verlag, Heidelberg. Germany.
22. **Hardman, J. K.** 1962. γ-Hydroxybutyrate dehydrogenase from *Clostridium aminobutyricum*. Methods Enzymol. 5:778–783.
23. **Hempel, J., K. Harper, and R. Lindahl.** 1989. Inducible (class 3) aldehyde dehydrogenase from rat hepatocellular carcinome and 2,3,7,8-tetrachlorodibenzo-*p*-dioxin-treated liver: distant relationship to the class 1 and 2 enzymes from mammalian liver cytosol/mitochondria. Biochemistry 28:1160–1167.
24. **Hidalgo, E., Y.-M. Chen, E. C. C. Lin, and J. Aguilar.** 1991. Molecular cloning and DNA sequencing of the *Escherichia coli* K-12 *ald* gene encoding aldehyde dehydrogenase. J. Bacteriol. 173:6118–6123.
25. **Jakoby, W. B.** 1963. Aldehyde dehydrogenases, p. 203–221. *In* P. D. Boyer (ed.), The enzymes, 2nd ed., vol. 7. Academic Press, New York.
26. **Jörnvall, H., B. Persson, and J. Jeffery.** 1987. Characteristics of alcohol/polyol dehydrogenases. The zinc-containing long-chain alcohol-dehydrogenases. Eur. J. Biochem. 167:195–201.
27. **Kenealy, W. R., and D. M. Waselefsky.** 1985. Studies on the substrate range of *Clostridium kluyveri*; the use of propanol and succinate. Arch. Microbiol. 141:187–194.
28. **Kessler, D., I. Leibrecht, and J. Knappe.** 1991. Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of *Escherichia coli* reside on a polymeric protein particle encoded by *adhE*. FEBS Lett. 281:59–63.
29. **Kletzin, A.** 1992. Molecular characterization of the *sor* gene, which encodes the sulfur oxygenase/reductase of the thermoacidophilic archaeum *Desulfurolobus ambivalens*. J. Bacteriol. 174:5854–5859.
30. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. Mol. Biol. 157:105–132.
31. **Lee, F.-J. S., L.-W. Lin, and J. A. Smith.** 1990. A glucose-repressible gene encodes acetyl-CoA hydrolase from *Saccharomyces cerevisiae*. J. Biol. Chem. 265:7413–7418.
32. **Lin, T., and W. A. Bridger.** 1992. Sequence of a cDNA clone encoding pig heart mitochondrial CoA transferase. J. Biol. Chem. 267:975–978.
33. **Lurz, R., F. Mayer, and G. Gottschalk.** 1979. Electron microscopic study on the quarternary structure of the isolated particulate alcohol-acetaldehyde dehydrogenase complex and on its identity with the polygonal bodies of *Clostridium kluyveri*. Arch. Microbiol. 120:255–262.
34. **Marathe, S., I. F. Connerton, and J. R. S. Fincham.** 1990. Duplication-induced mutation of a new *Neurospora* gene required for acetate utilization: properties of the mutant and predicted amino acid sequence of the protein product. Mol. Cell. Biol. 10:2638–2644.
35. **Nair, R. V., G. N. Bennett, and E. T. Papoutsakis.** 1994. Molecular characterization of an aldehyde/alcohol dehydrogenase gene from *C. acetobutylicum* ATCC 824. J. Bacteriol. 176:871–885.
36. **Palosaari, N. R., and P. Rogers.** 1988. Purification and properties of the inducible coenzyme A-linked butyraldehyde dehydrogenase from *Clostridium acetobutylicum*. J. Bacteriol. 170:2971–2976.
37. **Parales, R. E., and C. S. Harwood.** 1992. Characterization of the genes encoding β-ketoadipate:succinyl-coenzyme A transferase in *Pseudomonas putida*. J. Bacteriol. 174:4657–4666.
38. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85:2444–2448.
39. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. 55:339–372.
40. **Reid, M. F., and C. A. Fewson.** 1994. Molecular characterization of microbial alcohol dehydrogenases. Crit. Rev. Microbiol. 20:13–35.
41. **Saito, H., and K. I. Miura.** 1963. Preparation of transforming deoxyribonucleic acid by phenol treatment. Biochim. Biophys. Acta 72:619–629.
42. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
44. **Scherf, U., and W. Buckel.** 1991. Purification and properties of 4-hydroxybutyrate coenzyme A transferase from *Clostridium aminobutyricum*. Appl. Environ. Microbiol. 57:2699–2702.
45. **Scherf, U., B. Söhling, G. Gottschalk, D. Linder, and W. Buckel.** 1994. Succinate-ethanol fermentation in *Clostridium kluyveri*. Purification and characterization of 4-hydroxybutyryl-CoA dehydratase/vinylacetyl-CoA Δ^3-Δ^2-isomerase. Arch. Microbiol. 161:239–245.
46. **Schoberth, S., and G. Gottschalk.** 1969. Considerations on the energy metabolism of *Clostridium kluyveri*. Arch. Microbiol. 65:318–328.
47. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA 71:1342–1346.
48. **Smith, L. T., and N. O. Kaplan.** 1980. Purification, properties, and kinetic mechanism of coenzyme A-linked aldehyde dehydrogenase from *Clostridium kluyveri*. Arch. Biochem. Biophys. 203:663–675.
49. **Söhling, B., and G. Gottschalk.** 1993. Purification and characterization of a coenzyme-A-dependent succinate-semialdehyde dehydrogenase from *Clostridium kluyveri*. Eur. J. Biochem. 212:121–127.
50. **Thauer, R. K., K. Jungermann, H. Henninger, J. Wenning, and K. Decker.** 1968. The energy metabolism of *Clostridium kluyveri*. Eur. J. Biochem. 4:173–180.
51. **Walter, K. A., G. N. Bennett, and E. T. Papoutsakis.** 1992. Molecular characterization of two *Clostridium acetobutylicum* ATCC 824 butanol dehydrogenase isozyme genes. J. Bacteriol. 174:7149–7158.
52. **Weiser, J. N., J. M. Love, and E. R. Moxon.** 1989. The molecular mechanism of phase variation of *H. influenzae* lipopolysaccharide. Cell 59:657–665.
53. **Wierenga, R. K., P. Terpstra, and W. G. J. Hol.** 1986. Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. J. Mol. Biol. 187:101–107.
54. **Williamson, V. M., and C. E. Paquin.** 1987. Homology of the *Saccharomyces cerevisiae* ADH4 to an iron-activated alcohol dehydrogenase from *Zymomonas mobilis*. Mol. Gen. Genet. 209:374–381.
55. **Wolff, R. A., G. W. Urben, S. M. O'Herrin, and W. R. Kenealy.** 1993. Dehydrogenases involved in the conversion of succinate to 4-hydroxybutanoate by *Clostridium kluyveri*. Appl. Environ. Microbiol. 59:1876–1882.
56. **Yan, R.-T., and J.-S. Chen.** 1990. Coenzyme A-acylating aldehyde dehydro-

genase from *Clostridium beijerinckii* NRRL B592. Appl. Environ. Microbiol. 56:2591–2599.
57. **Yang, W., E. Li, T. Kairong, and S. L. Stanley, Jr.** 1994. *Entamoeba histolytica* has an alcohol dehydrogenase homologous to the multifunctional *adhE* gene product of *Escherichia coli*. Mol. Biochem. Parasitol. 64:253–260.
58. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119.
59. **Yoon, K. H., and M. Y. Pack.** 1990. Nucleotide sequence of the *Zymomonas mobilis* alcohol dehydrogenase II gene. Nucleic Acids Res. 18:187.
60. **Young, M., N. P. Minton, and W. L. Staudenbauer.** 1989. Recent advances in the genetics of the clostridia. FEMS Microbiol. Rev. 63:301–326.
61. **Youngleson, J. S., W. A. Jones, D. T. Jones, and D. R. Woods.** 1989. Molecular analysis and nucleotide sequence of the *adh1* gene encoding an NADPH-dependent butanol dehydrogenase in the gram-positive anaerobe *Clostridium acetobutylicum*. Gene 78:355–364.